

Study and Implementation of some Quantitative Trading Models

Author: Emiliano Sánchez López

Advisor: Josep J. Masdemont

Universitat Politècnica de Catalunya
Facultat de Matemàtiques i Estadística

Master Thesis

**Study and Implementation of
some Quantitative Trading
Models**

Emiliano Sánchez López

Advisor: Josep J. Masdemont

Matemàtica Aplicada I

Preface

This Master of Science Thesis is due to a collaboration agreement between the Centre de Recerca Matemàtica of Catalonia (CRM), Universitat Politècnica de Catalunya (UPC) and Everis (IT-Consulting).

The project was supported by a Master Thesis Grant 2010 awarded by the CRM, which its objective was to promote and create technology transfer projects, specially within the applied mathematics field, between local companies and universities.

The research project deals with mathematical and statistical methods for deployment of quantitative models for financial investments in a high-frequency setting. The algorithms are written in Matlab[®] and tested with minutely data of the Dow Jones Industrial Average.

I would like to thank my supervisor, Professor Josep J. Masdemont, for spending his precious time carefully reading and discussing this work. Also I would like to thank him for proposing me for this grant in the first place.

I would also to express my gratitude to Karel Komorád at Everis for spending several hours of his working time discussing and improving with me this project.

Last but not least I would like to say a very special thank you, to my parents, my sister and brothers and to Colette Roitsch for all the support and patience given to me while writing this thesis.

This thesis was written according to the guidelines presented by the Faculty of Mathematics and Statistics of UPC.

Emiliano Sánchez
Barcelona, Spain
June, 2011

Abstract

Quantitative or algorithmic trading is the automatization of investments decisions obeying a fixed or dynamic sets of rules to determine trading orders. It has increasingly made its way up to 70% of the trading volume of one of the biggest financial markets such as the New York Stock Exchange (NYSE). However, there is not a significant amount of academic literature devoted to it due to the private nature of investment banks and hedge funds.

This projects aims to review the literature and discuss the models available in a subject that publications are scarce and infrequently. We review the basic and fundamental mathematical concepts needed for modeling financial markets such as: stochastic processes, stochastic integration and basic models for prices and spreads dynamics necessary for building quantitative strategies. We also contrast these models with real market data with minutely sampling frequency from the Dow Jones Industrial Average (DJIA).

Quantitative strategies try to exploit two types of behavior: trend following or mean reversion. The former is grouped in the so-called technical models and the later in the so-called pairs trading. Technical models have been discarded by financial theoreticians but we show that they can be properly cast into a well defined scientific predictor if the signal generated by them pass the test of being a Markov time. That is, we can tell if the signal has occurred or not by examining the information up to the current time; or more technically, if the event is \mathcal{F}_t -measurable. On the other hand the concept of pairs trading or market neutral strategy is fairly simple. However it can be cast in a variety of mathematical models ranging from a method based on a simple euclidean distance, in a co-integration framework or involving stochastic differential equations such as the well-known Ornstein-Uhlenbeck mean reversal ODE and its variations.

A model for forecasting any economic or financial magnitude could be properly defined with scientific rigor but it could also lack of any economical value and be considered useless from a practical point of view.

This is why this project could not be complete without a backtesting of the mentioned strategies.

Conducting a useful and realistic backtesting is by no means a trivial exercise since the “laws” that govern financial markets are constantly evolving in time. This is the reason because we make emphasis in the calibration process of the strategies’ parameters to adapt the given market conditions.

We find out that the parameters from technical models are more volatile than their counterpart from market neutral strategies and calibration must be done in a high-frequency sampling manner to constantly track the currently market situation.

As a whole, the goal of this project is to provide an overview of a quantitative approach to investment reviewing basic strategies and illustrating them by means of a back-testing with real financial market data.

The sources of the data used in this project are Bloomberg for intra-day time series and Yahoo! for daily prices. All numeric computations and graphics used and shown in this project were implemented in MATLAB[®] scratch from scratch as a part of this thesis. No other mathematical or statistical software was used.

Key words: stochastic ordinary differential equations, diffusion processes, jump processes, stochastic models, statistical methods, econometrics, optimal timing strategies, high-frequency data, pairs trading, co-integration, portfolio choice, investment decisions.

MSC2000: 60H10, 60J60, 60J75, 91B70, 91G70.

JEL: C41, C58, G11, G12, G14.

Contents

List of Figures	iii
List of Tables	v
Chapter 1. Introduction and Background	1
1. Motivation	1
2. Objectives and structure of the project	2
3. The principle of absence of arbitrage	3
4. Literature review	4
5. Theoretical background	5
6. Empirical analysis of financial time series	18
7. Filtering in finance	33
8. Statistical testing of trading strategies	38
9. Performance measurements of financial investments	39
Chapter 2. Technical Trading	45
1. Moving average crossings indicators	46
2. Trend crossing methods	47
3. Technical oscillators	48
4. Patterns in financial time series	49
5. Moving average convergence divergence	49
6. Mixed indicators models	50
7. Finite Markov times	50
Chapter 3. Pairs Trading	53
1. The distance method between vector prices	54
2. Co-integration	56
3. Stochastic spread	62
4. Stochastic residual spread	69
Chapter 4. Numerical Results	71
1. Testing a technical trading model	71
2. Testing pairs trading models	73
Chapter 5. Conclusions and Future Work	83
References	87

Glossary	91
Index	93
Appendix A. DJIA Companies	95

List of Figures

1	A (cumulative) realization of a Wiener process.	9
2	A realization of a Wiener process in the xy-plane.	10
3	A realization of a Wiener process in the xyz-space.	11
4	A realization of a GBM process with $x_0 = 10$, $\mu = 0.25$ and $\sigma = 0.3$.	14
5	A realization of a Jump-Diffusion process with $x_0 = 10$, $\mu = 0.25$, $\sigma = 0.3$, $a = 0$, $b = 0.15$ and $\lambda = 3.45$.	16
6	A realization of an Ornstein-Uhlenbeck process with $\mu = 0$, $\theta = 2$ and $\sigma = 1$.	17
7	Probability distributions of XOM log-returns (left). Probability of negative return measured as a function of the time interval Δt (right). No scaling behavior is detected.	21
8	Scaling law for XOM. The time interval goes from 1 min to 60 min.	23
9	Intraday autocorrelation function for XOM.	25
10	Intraday volatility profile for the Dow Jones Index.	25
11	Intraday autocorrelation function for r_t^2 XOM during a trading week.	26
12	Variance plot for XOM gives the estimate for the long memory parameter β .	26
13	Probability density function for the correlation coefficients of the DJIA: $P(\rho)$ for each of the five calendar years from 2006 to 2010.	28
14	Taxonomy of the DJIA portfolio. Polar Hierarchical tree.	30
15	Taxonomy of the DJIA portfolio. Minimum spanning tree obtained using the Kruskal algorithm.	31
16	Empirical distribution of the eigenvalues of a pure random matrix (left). Limiting distributions of the eigenvalues for a random matrix (right).	32

17	Histogram of eigenvalues of \mathbf{C} , where the correlation matrix is extracted from the DJIA and the theoretical density for $Q=5$. Inset: same plot, but including the highest eigenvalue corresponding to the ‘market’, which is found to be ~ 9 times higher than λ_{\max} .	33
18	Full Kelly investor choose $f=1$ which is equivalent to maximize the log-utility function $u(x) = \log(c)$. Partial Kelly strategies ($f < 1$) are equivalent to maximize $u(x; \alpha) = \alpha \cdot x^\alpha$, with $\alpha < 1$	43
19	Moving average crossing for General Electric (upper panel). Trading volume (lower panel).	47
20	Standard deviation rule (upper panel). α -quantile rule (lower panel). The squares represent entry points in long positions, conversely the stars represent the exit points.	55
21	Log prices of AT&T and Verizon.	63
22	Spread of AT&T and Verizon. Inset: A spread roundtrip from 2-jul-2010 to 2-ago-2010.	64
23	Geometrical interpretation of the deterministic drift of the O-U process.	68
24	Sharpe’s ratio iso-surface for a MA/RSI implementation for XOM. The darker areas represents higher SR.	72
25	P&L (out-of-the-sample) generated by a mixed model calibrated with $\Delta t = 70$, MA=160 and RSI=140.	73
26	T and VZ spread (in-sample) with rolling standard deviation (upper panel). Position vector for the first asset (lower panel).	75
27	P&L generated (out-of-the-sample) by the distance method with the standard deviation rule.	76
28	Heat Map: ADF test for the DJIA price matrix. The darker squares represents likelihood of co-integrability.	77
29	Sharpe ratio surface for the co-integration method (in-sample).	78
30	P&L generated (out-of-the-sample) by the co-integration method for HPQ and DIS.	79
31	Sharpe Ratio Surface (in-sample) for the stochastic spread method.	80
32	P&L generated (out-of-the-sample) by the stochastic spread method for T and VZ.	81
33	A sketch of a quantitative trading system.	84

List of Tables

1	Moments of logarithmic return distribution for the most traded securities in the Dow Jones index in the period 01-Jun-2010 to 08-Nov-2010.	20
2	Estimated tail indices for four Dow Jones stocks with $m=150$.	22
3	Drifts exponents for U.S securities.	23
4	The observed maximum and minimum values of $\rho_{i,j}$ for the DJIA portfolio for each of the five calendar years from 2006 to 2010.	27
5	Recursive equations for the KF.	66
6	Optimal parameters calibrated in the in-sample data set (75% of the data points).	73
7	Performance (out-of-the-sample) of the MA/RSI method calibrated with $\Delta t = 70$, MA=160 and RSI=140 for XOM.	74
8	Performance (out-of-the-sample) of the distance method for T and VZ calibrated with a rolling window of 140 minutes.	75
9	Performance (out-of-the-sample) of the co-integration method calibrated with a rolling window of 250 minutes for HPQ and DIS.	78
10	Performance (out-of-the-sample) of the stochastic spread method calibrated with a rolling window of 25 minutes and updated every 5 minutes for T and VZ.	80

Notation and Abbreviations

$d_{i,j}$	Euclidean distance between asset i and j .
\mathbf{C}	Empirical correlation matrix.
\mathbb{E}	Expectation operator.
\mathcal{F}	Filtration.
Φ	Risk measure.
$\hat{\gamma}$	Hill estimator.
J_t	Jump process.
K	Kalman gain.
λ_i	Eigenvalue i of the correlation matrix.
\mathbb{P}	Probability measure.
r_t	Logarithmic return.
$\rho_{i,j}$	Correlation coefficient between asset i and j .
σ_i	Volatility of asset i .
τ	Markov time.
\mathbb{V}	Variance operator.
W_t	Wiener process, Brownian motion.
AR	Auto-regressive process.
DAO	Deterministic arbitrage opportunity.
DJIA	Dow Jones Industrial Average.
EFK	Extended Kalman filter.
ES	Expected shortfall.
EM	Expectation-Maximization.
GARCH	Generalized auto-regressive conditional heteroscedastic.
GBM	Geometric Brownian motion.
KF	Kalman filter.
MA	Moving average.
MM	Modigliani-Modigliani indicator.
MST	Minimum spanning tree.

Notation and Abbreviations

NYSE	New York Stock Exchange.
ODE	Ordinary differential equation.
OU	Ornstein Uhlenbeck process.
P&L	Profit and loss.
RMT	Random matrix theory.
SAO	Statistical arbitrage opportunity.
SDE	Stochastic differential equation.
SR	Sharpe ratio.
TA	Technical analysis.
UKF	Unscented Kalman filter.
VaR	Value-at-risk.
VAR	Vector autoregressive.

Chapter 1

Introduction and Background

1. Motivation

Statistical arbitrage is a long horizon trading strategy that generates a risk-free profit. It is based on taking advantage of the statistical mispricing of one or more assets and expecting its convergence towards its fair value. The concept of statistical arbitrage is motivated by numerous empirical studies [17, 39, 42, 41] that construct trading strategies to profit from persistent anomalies and contradicting market efficiency, a building block in financial theory, first proposed by Fama [15].

The motivation to pursue a better understanding of statistical arbitrage is twofold. From a theoretical point of view, we address a number of questions in order to have a better understanding of the characteristics of the stochastic process that drives the securities involved. From a practical point of view, it is well known that [statistical arbitrage](#) strategies are common among many hedge funds, investment banks and other market participants. There are estimates that algorithmic trading is responsible for as much as the 70% of the trading volume in the U.S in 2009¹. However there is not a significant amount of academic literature devoted to it due to its proprietary nature.

A common approach to statistical arbitrage would be the following. Construct a stationary, [mean reverting](#) synthetic asset by means of a linear combination of securities. The trades are entered when the process reaches an extreme value, and exited when the process reverts to some mean value.

¹“The Impact of High Frequency Trading: Manipulation, Distortion or a Better-Functioning Market”, Financial Times Press, November 9, 2009.

2. Objectives and structure of the project

This project aims to review the literature and discuss the models available in a subject that publications are scarce and infrequently. We want to review, and up to certain extent, analyze the concepts needed for modeling financial markets and for building quantitative strategies driven by mathematical and statistical methods. We also want to backtest these models with real market data.

The **first chapter** could be seen as a primer for some background of both theoretical and empirical methods frequently used in the field of mathematical finance. Here we have included definitions of the concept of [arbitrage](#) and statistical arbitrage, we refresh concepts from probability, stochastic processes and stochastic integration needed for the discussion of models for prices and spreads dynamics. These ideas, concepts and techniques are contrasted in an empirical analysis conducted with a minutely high frequency data set. We also discuss some filtering techniques in order to obtain better estimates from some non-observable time series and finally we discuss a statistical approach to test the strategies and discuss how to measure the reward and risk for describing the wealth dynamics generated by the quantitative strategies reviewed in the subsequent chapters.

Equipped with chapter [1](#) we are able to discuss two types of trading strategies. Namely, [technical trading](#) models and pairs trading models. In **chapter 2** we deal with the question whether technical trading (widely used by market participants) can be formalized or lacks of scientific value. In **chapter 3** we move into one of the most used market neutral strategy, the so-called pairs trading. Although the basic concept of this strategy remains the same, there are different versions of the same idea from the most basic non-parametric “distance method”, going through an econometric approach with the “co-integration method” and finally in a continuous time setting the “stochastic spread method” and its variations.

After reviewing in detail the available models in the previous chapters, we proceed to discuss them numerically in **chapter 4**. For this purpose one can proceed with two different approaches. The first one is to study the behavior of the models by simulating artificial data sets by manipulating the parameters such as drifts and volatilities, in the case of price dynamics, and level and speed of mean reversion, in the case of spread dynamics in order to compare the different outcomes. The second one, and the most important from a practical point of view, is to test models with real market data. This is done in **chapter 4** in a high-frequency setting involving data from stocks of the Dow Jones Industrial Average (DJIA) Index.

3. The principle of absence of arbitrage

The *principle of absence of arbitrage* is perhaps the most fundamental principle of finance theory.

Although financial market modelers are equipped with a rich toolbox (stochastic calculus, optimization, advance probability theory, etc.) they do not yet have the right laws of science to exploit such as mechanical or electrical engineers do with Newton's laws or Maxwell's equations respectively [9].

That is, most models only rely on the law of one price or no arbitrage², which states that if the payoff of an asset can be synthetically replicated by a portfolio of assets, the price of the portfolio and the price of the asset it replicates, must be the same.

Definition 1. A *deterministic arbitrage opportunity (DAO)* is a zero-cost trading strategy that offers the possibility of a gain with no possibility of loss.

The fundamental theorem of the financial theory [21] establish a link between the absence of DAO and the existence of a positive pricing kernel.³ An arbitrage possibility is thus essentially equivalent to the possibility of making a positive amount of money out of nothing without taking any risk. In financial market parlance this is called “a free lunch”. A *statistical arbitrage opportunity (SAO)* [3] is a trading strategy in which the expected payoff is positive and the conditional expected payoff in each final state of a given economy is nonnegative.

The main difference between a SAO and a DAO, is that the SAO strategy admits negative payoffs at any time. The definition of statistical arbitrage is independent of any equilibrium model or formulation of expected returns, and its existence contradicts market efficiency.

Let $\{v_i\}_{i=1,\dots,n}$ be a sequence of discounted portfolio values generated by a self-financing strategy (funds are neither added to nor withdrawn for). We denote Δv the increment of trading strategy and $v(n)$ its terminal value. A statistical arbitrage requires the trading profits of a zero cost, self-financing trading strategy to satisfy four axioms:

Definition 2. A *statistical arbitrage opportunity* is a zero initial cost, self-financing trading strategy with cumulative discounted trading profits $v(n)$ and incremental discounted trading profits $\Delta v(n)$ such that:

²One can also consider the first axiom of financial mathematics: the time value of money, and some useful principles such as the benefits of diversification and value of the right to choose.

³Also called stochastic discount factors that are used to represent valuation operators in dynamic stochastic economies.

- (1) $v(0) = 0$
- (2) $\lim_{n \rightarrow \infty} \mathbb{E}^P[v(n) > 0]$
- (3) $\lim_{n \rightarrow \infty} \mathbb{P}(v(n) < 0) = 0$
- (4) $\lim_{n \rightarrow \infty} \mathbb{V}[\Delta v(n) | \Delta v(n) < 0] = 0$

These axioms indicate that the portfolio: (1) is [self-financing](#), (2) in the limit has positive expected discounted profits, (3) has a probability of a loss converging to zero (4) as well as semi-variance convergence to zero. The fourth axiom is a consequence of the third. A given investor is only concerned with the variability of the trading strategy below zero (semi-variance). Because in the limit the probability of the strategy being negative is zero, also it is the semi-variance [\[24\]](#).

4. Literature review

Although the first appearance (in financial industry) of quantitative trading models that prescribe investment decisions dates back to the mid-1980 [\[48\]](#), these models did not caught immediately the attention of the financial academic community. Nevertheless, the early literature started to focus on the study of simple technical indicators for equity markets. Some remarkable examples are Brock et al. [\[7\]](#) or LeBaron [\[30, 31\]](#) which demonstrate the statistical significance of technical trading rules against well-known null models such as the random walk, the AR(1) and the GARCH-M.

The study of Neftci [\[40\]](#) investigates the statistical properties of technical analysis in order to determine if there is any scientific basis for the rules developed by practitioners without any reference to any formalism. Also the author deals with a comparison to the Wiener-Kolmogorov prediction theory, and proves that when the underlying processes are nonlinear, the trading rules outperform the mentioned theory.

Overall, the scope of the most recent literature supports technical models, but it is restricted to univariate technical rules. Besides the trading rules inspired by technical analysis, exists another very popular technique that exploits the relationship of time series of order > 1 ; the so-called “[pairs trading](#)”. In its most common form, pairs trading involves forming a portfolio of two related stocks whose relative pricing is away from long-run state. When the spread between them widens, one must go long on the relatively undervalue stock and short on the relatively overvalue stock, if history repeats itself, prices will converge and the arbitrageur will lock-in a profit.

The most referenced work to a nonparametric approach to pairs trading includes Gatev et al. [\[17\]](#), Nath [\[39\]](#) and Perlin [\[42\]](#). The first paper

examine empirically the U.S equity market, it proposes to minimize the euclidean distance of the time series and to trade upon a simple standard deviation rule strategy. They show that this strategy is profitable after including trading costs. The second research paper deals with the same strategy but with government debt. It calculates the [spread](#) between bonds and its empirical distribution. Then it trades upon a certain quantile of the empirical distribution. The last of these three authors study the same strategy in the Brazilian financial market. His conclusions concur with the previous research studies. Perlin also extends the same strategy to a multivariate case in [\[41\]](#).

Vidyamurthy [\[48\]](#), implements pairs trading based on a [co-integration](#) framework, without empirical results. Herlemont [\[22\]](#) discuss a co-integration approach to trading from a theoretical point of view with emphasis in unit root testing and also reviews a multivariate formulation, the vector error correction model. Trapletti et al. [\[47\]](#) present a co-integration analysis triangle of foreign exchange currencies in a high frequency setting. They evaluate empirically the out-of-sample forecasting power of a VAR and show that results are economically significant even when including transaction costs.

Elliott et al. [\[12\]](#) propose an analytical framework in a continuous time setting to model the spread between two assets with a mean-reverting Gaussian Markov chain model. They Also apply a Kalman filter to estimate a parametric model of the spread. Do et al. [\[11\]](#) analyze the existing methods in detail and propose a new methodology to model mispricing for pairs trading. They take into account an asset pricing relationship, conversely to the majority of methodologies, purely based on statistical considerations. Meucci [\[36\]](#) provides an example of multivariate Ornstein-Uhlenbeck process application to swap data and gives a geometrical interpretation of mean reverting dynamics.

5. Theoretical background

5.1. Probability models in financial mathematics

Essentially, a probability model consists on a filtered probability space where the variables of interest are defined. A filtered probability space is made of a sample space of elementary events, a sigma-algebra of events, a probability defined on that sigma-algebra and a filtration of increasing sigma-algebra.

Consider a single stock price S_t at discrete time $t=1,2,\dots,T$. Denote by Ω the set of all possible values of the stock during these times

$$\Omega = \{\omega : \omega = (S_1, S_2, \dots, S_T)\}$$

To model uncertainty about the price in the future, we “list” all possible future prices, and we call them *possible states of the world*. As time passes, more and more information is revealed about the true state of the world. Thus the true state of the world is contained in smaller sets, subsets of Ω .

The information available to investors at time t , \mathcal{F}_t , is the history of stock prices up to time t . If we assume that the stock price can go up by a factor u or down by a factor d , then the relevant information reduces to the knowledge of the movements at each time

$$\Omega = \{\omega : \omega = (a_1, a_2, \dots, a_T)\}, \quad a_t = u \text{ or } d.$$

Consider a two-period model. At $t=0$ we have no information about S_1 and S_2 , so $\mathcal{F}_0 = \{\emptyset, \Omega\}$. All we know about the true state of the world is in Ω . Suppose at $t=1$ that the stock went up. Then we know that the true state of the world lies in A , and not inside its complement \bar{A} , where

$$A = \{(u, S_2), S_2 = u \text{ or } d\} = \{(u, u), (u, d)\}.$$

Thus our information at time $t=1$ is,

$$\mathcal{F}_1 = \{\emptyset, \Omega, A, \bar{A}\}.$$

Since we don't lose information, $\mathcal{F}_0 \subset \mathcal{F}_1$.

\mathcal{F}_t is called an algebra sets. \mathcal{F} is a field if

- (1) $\emptyset, \Omega \in \mathcal{F}$.
- (2) $A \in \mathcal{F}, \bar{A} \in \mathcal{F}$.
- (3) If $A \in \mathcal{F}$, and $B \in \mathcal{F}$ then $A \cup B \in \mathcal{F}$.

A collection of fields is called a filtration

$$\mathbb{F} = \{\mathcal{F}_0, \mathcal{F}_1, \dots, \mathcal{F}_t, \dots, \mathcal{F}_T\}, \quad \mathcal{F}_t \subset \mathcal{F}_{t+1}.$$

\mathbb{F} is used to model a flow of information. As time passes, an observer knows more and more detailed information, represented by finer and finer partitions of Ω .

Definition 3. A probability space consist of a triple (Ω, \mathcal{F}, P) where

- (1) Ω is a space of points ω , called the sample space and sample points.
- (2) \mathcal{F} is a σ -algebra of subsets of Ω . These subsets are called events.
- (3) $P(\cdot)$ is a probability measure on \mathcal{F} .

Definition 4. A function X on Ω is called \mathcal{F} -measurable or random variable on (Ω, \mathcal{F}) if all the sets $\{X = x_i\}, i = 1, \dots, k$ are members of \mathcal{F} .

Definition 5. A countable stochastic process is a sequence of random variables X_1, X_2, \dots defined on a common probability space (Ω, \mathcal{F}, P) . A stochastic process is called adapted to filtration \mathbb{F} if for all $t = 0, 1, \dots, T$, X_t is \mathcal{F} -measurable.

Definition 6. τ is called a random time or Markov time if its is a non-negative random variable, which can also take value ∞ on (Ω, \mathcal{F}_T) . Given the filtration $\mathbb{F} = \{\mathcal{F}_0, \mathcal{F}_1, \dots, \mathcal{F}_t, \dots, \mathcal{F}_T\}$, τ is called stopping or Markov time with respect to this filtration if for each $t = 0, 1, \dots, T$ the event

$$\{\tau < t\} \in \mathcal{F}_t.$$

By observing the information contained in \mathcal{F}_t we can decide whether the event has occurred or not.

So far we have reviewed some probability concepts in a discrete time setting, now extend them for a continuous sample space.

Definition 7. A σ -algebra is an algebra, which is closed with respect to countable unions and countable intersections of its members, that is a collection of subsets of Ω that satisfies

- (1) $\emptyset, \Omega \in \mathcal{F}$.
- (2) $A \in \mathcal{F} \Rightarrow \bar{A} \in \mathcal{F}$
- (3) $A_1, A_2, \dots, A_n, \dots \in \mathcal{F}$ then $\cup_{n=1}^{\infty} A_n \in \mathcal{F}$ (and also $\cap_{n=1}^{\infty} A_n \in \mathcal{F}$).

Any subset B of Ω that belongs to \mathcal{F} is called a measurable set.

Definition 8. Strict Stationarity

A process X_t defined in a probability space (Ω, \mathcal{F}, P) is stationary in the strict sense if, for any $\{t_1, \dots, t_k\}$ where $k = 1, 2, \dots$

$$F_{t_1, \dots, t_k} = F_{t_1+h, \dots, t_k+h}$$

for all h , where F_{t_1, \dots, t_k} is the joint distribution function of $(X_{t_1}, \dots, X_{t_k})$.

Definition 9. Weak Stationarity

A process X_t defined in a probability space (Ω, \mathcal{F}, P) is stationary in the weak sense if it is square integrable,

$$\mathbb{E}(X_t) = \mu$$

and

$$\text{cov}(X_t, X_s) = f(|t - s|).$$

This implies that $\text{var}(X_t) = \text{cov}(X_t, X_t) = f(0)$ is also constant in t . In the Gaussian case, since the process is completely specified by its mean and covariance functions, weak and strong stationarity are equivalent.

Definition 10. *Co-Integration*

Two nonstationary time series A_t and B_t are co-integrated if there is some $\gamma \in \mathbb{R}$ such that $A_t - \gamma B_t$ is stationary in the weak sense.

Definition 11. *Ergodicity*

A d -dimensional random process, $\{\mathbf{X}_t, t \geq 0; \mathbf{X}_0 = \mathbf{x}\}$ is ergodic if there is some measure μ such that for any μ -integrable function f ,

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t f(\mathbf{X}_s) ds = \int_{\mathbb{R}^d} f d\mu.$$

Intuitively this means that the long run average behavior mimics the instantaneous behavior. So an ergodic process will visit each state infinitely often, and this implies that the time between visits is certainly finite.

5.2. Stochastic processes in financial mathematics

A stochastic process is an umbrella term for any collection of random variables $\{X_t\}$ depending on time. In practice, we typically observe only a single realization of this process, a simple path, out of a multitude of possible paths. Brownian motion is the main process in the calculus of continuous process and Poisson jumps is the main process in the calculus of process with discontinuities. Here we are only going to deal with the former.

R. Brown was the first to describe the irregular and random motion of a pollen particle suspended in fluid in 1828. L. Bachelier was the first to propose an approach based on Brownian motion to model stock prices in his Ph.D. thesis “*Théorie de la spéculation*” [2]. A. Einstein, independently, in 1905 derived the equations of Brownian motion arguing that the movement is due to bombardment of the particle by molecules of the fluid. N. Wiener was the first to put this phenomena in a solid mathematical ground in his famous paper “*Differential Space*” [49]. The names Brownian motion or Wiener process are interchangeable. The exposition in this section comes from [29].

A Wiener process $\{W_t\}$ is a stochastic process with the following properties:

- (1) *Continuity of paths.* $W_t, t \geq 0$ are continuous functions of t .
- (2) *Normal increments.* $W_t - W_s$, has a normal distribution with zero mean and variance $t-s$.
- (3) *Independence of increments.* $W_t - W_s$, for $t > s$, is independent of the past, that is, of $W_u, 0 \leq u \leq s$ or the σ -algebra generated by $W_u, u \leq s$.

An occurrence of Wiener process observed from time 0 to time T , is a random function of t on the interval $[0, T]$. It is called realization, a path or trajectory.

The realization of $\{W_t\}$ has the following properties with probability one:

- (1) Is not monotone in any interval, no matter how small the interval is.
- (2) Is not differentiable at any point.
- (3) Has infinite variation on any interval, no matter how small it is.
- (4) For any t has a quadratic variation on $[0, t]$.

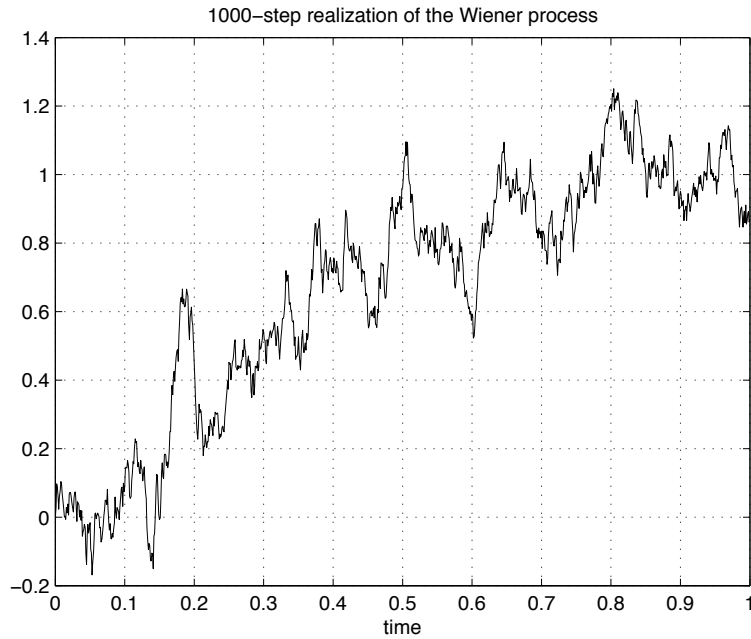


FIG. 1. A (cumulative) realization of a Wiener process.

A positive quadratic variation implies infinite variation, so property 3 follows from property 4. Since a monotone function has finite variation, property 1 follows from property 3. A continuous function with a

bounded derivative is of finite variation. Therefore it follows from property 3 that $\{W_t\}$ can not have a bounded derivative on any interval, no matter how small the interval is.

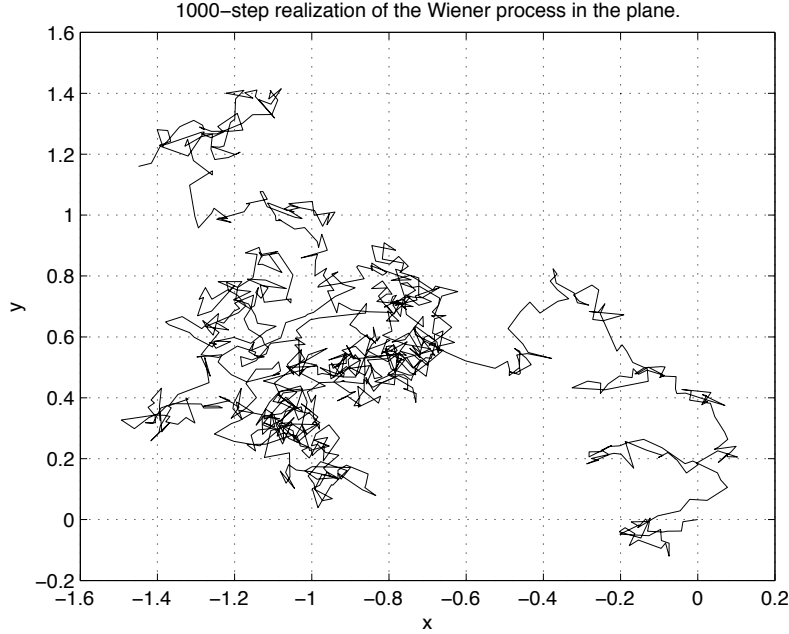


FIG. 2. A realization of a Wiener process in the xy-plane.

Another important characteristic of the Brownian motion is that if we know the present state of the process, then the future behavior is independent of its past. This is called the *Markov property*. The process $\{W_t\}$ has the Markov property if the conditional distribution of $\{W_{t+s}\}$ does not depend on past values.

Definition 12. $\{W_t\}$ is a Markov process if for any t and $s > 0$, the conditional distribution of $\{W_{t+s}\}$ given \mathcal{F}_t is the same as the conditional distribution of $\{W_{t+s}\}$ given W_t , that is,

$$P(W_{t+s} \leq y | \mathcal{F}_t) = P(W_{t+s} \leq y | W_t), \quad a.s.^4$$

Definition 13. A Wiener process in dimension two or higher is a random vector $\mathbf{W}_t = (W_t^1, W_t^2, \dots, W_t^n)$ with all coordinates W_t^i being independent one-dimensional Wiener process.

Remark 1. In dimensions one and two Brownian motion is recurrent. That is, it will come back to a neighbourhood, however small, of any point. In dimensions three and higher, Brownian motion is transient.

⁴almost surely, that is, for all paths except a set of zero measure.

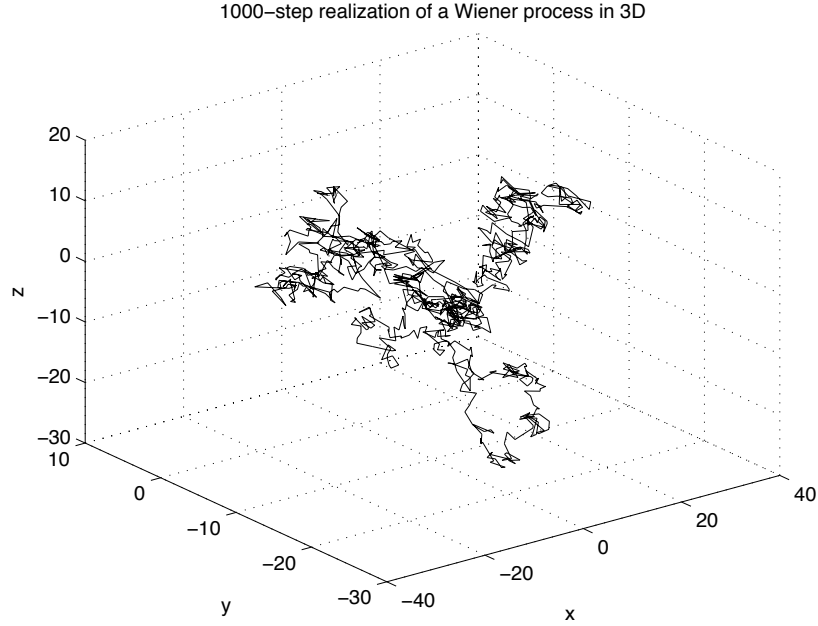


FIG. 3. A realization of a Wiener process in the xyz-space.

It will leave a ball, however large, around any point never to return [43].

5.3. Stochastic Integration

While a definite integral is a number and an indefinite integral is a function, a stochastic integral is a random variable if the time interval is fixed or a stochastic process if the time interval is variable.

The objective is to make sense of

$$\int_0^t b_s dW_s,$$

where $\{W_t\}_{t \geq 0}$ is the standard Wiener process and $\{b_t\}_{t \geq 0}$ is a process adapted to $\{W_t\}$.

If we were to define the Riemann integral $\int_0^T f_t dt$ then we would divide the interval $[0, T]$ into n subintervals (where n is a large positive integer) by choosing the points $t_0 = 0, t_1 = \frac{T}{n}, t_2 = \frac{2T}{n}, \dots, t_n = \frac{nT}{n} = T$. Then we would choose a point t_i^* in the subinterval $[t_i, t_{i+1}]$ and form the appropriate sum. If the limit of these sums as n goes to infinity exists and is independent of the choice of the t_i^* then the Riemann integral of our function f is defined by $\int_0^T f_t dt = \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} f(t_i^*)(t_{i+1} - t_i)$.

This definition generalizes to the case where dt is replaced by dg_t , where g_t is a differentiable function (or a function of finite variation) in the interval $[0, T]$. Then the integral is defined as follows $\int_0^T f_t dg_t = \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} f(t_i^*)(g(t_{i+1}) - g(t_i))$. This is the Riemann-Stieltjes integral of f .

However, the Wiener process W_t , as a function of t , is nowhere differentiable. Following a “Riemann-Stieltjes like” definition of a stochastic process will be t_i^* -dependent. Indeed, choosing for t_i^* the left end-point of each subinterval, yields a different sum than choosing the right end-point. The average values of the mentioned sums are different and therefore can not have the same limit. The above discussion suggests that in order to make sense of the stochastic integral one must choose t_i^* consistently. That is, the same position with respect to the end-points of each subinterval $[t_i, t_{i+1}]$. More precisely, $t_i^* = t_i + s(t_{i+1} - t_i)$, for some s with $0 \leq s \leq 1$. We have infinitely possibilities to choose s . The Itô integral uses $s = 0$.

Now the next step is to recall the characteristic function of an interval $b_t = \sum_{i=0}^{n-1} c_i \mathbb{1}_{(t_i, t_{i+1}]}$. Then by forming linear combinations of characteristic functions with random variables as coefficients we can construct the stochastic process.

So, the Itô integral is defined as the limiting sum

$$(1) \quad \int_0^t b_s dW_s = \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} c_i (W_{t_{i+1}} - W_{t_i}).$$

where $0 = t_0 < t_1, \dots, < t_n = T$ and c_i are random variables.

There are some relevant properties of the Itô integral that are useful in our context

$$(2) \quad \mathbb{E} \left[\int_0^t b_s dW_s \right] = 0.$$

$$(3) \quad \mathbb{E} \left[\left(\int_0^t b_s dW_s \right)^2 \right] = \int_0^t \mathbb{E}[b_s^2] ds.$$

Combining these observations gives that

$$(4) \quad \int_0^t b(s) dW_s \sim \mathcal{N}(0, \int_0^t b(s)^2 ds),$$

where $\mathcal{N}(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2 .

5.4. Models for price dynamics

The most common model to describe the dynamics of security prices is the geometric Brownian motion (GBM). Let us consider that the infinitesimal change of an asset x at time t is modeled by a stochastic differential equation (SDE),

$$(5) \quad dx_t = \mu x_t dt + \sigma x_t dW_t, \quad t \in [0, \infty).$$

Dividing by x_t and integrating over $(0, t)$ we get,

$$(6) \quad \int_0^t \frac{1}{x_s} dx_s = \int_0^t \mu dt + \int_0^t \sigma dW_s.$$

The first integral in the right-hand side does not contain any randomness. Since the coefficient of the second term is constant we can use (1) and the integral results in

$$(7) \quad \int_0^t \sigma dW_s = \sigma(W_t - W_0),$$

where $W_0 = 0$ by definition of the Wiener process. Thus,

$$(8) \quad \int_0^t \frac{1}{x_s} dx_s = \mu t + \sigma W_t.$$

Any solution of the SDE (5) must satisfy the above integral. Since the integrand in the left-hand gives us a hint, we introduce the following function $f(x_t, t) = \log(x_t)$.

Itô's lemma tells us that,

$$(9) \quad df(x_t, t) = \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial x_t} dx_t + \frac{1}{2} \frac{\partial^2 f}{\partial x_t^2} dx_t^2$$

Using (5) divided by x_t and the derivatives of $f(x_t, t)$ we get the total derivative,

$$(10) \quad df(x_t, t) = (\mu - \frac{1}{2}\sigma^2)dt + \sigma dW_t$$

Now, simply integrating over $(0, t)$ and exponentiating we find the solution of the SDE governed by a GBM

$$(11) \quad x_t = x_0 \cdot \exp\{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t\}$$

Remark 2. *The drift rate of $f(x_t, t) = \log(x_t)$ is less than μ , because $\log(x_t)$ is a concave function of x_t . So as x_t is uncertain, the expected value of $\log(x_t)$ changes by less than the logarithm of the expected value of x_t . This is a consequence of Jensen's inequality.*

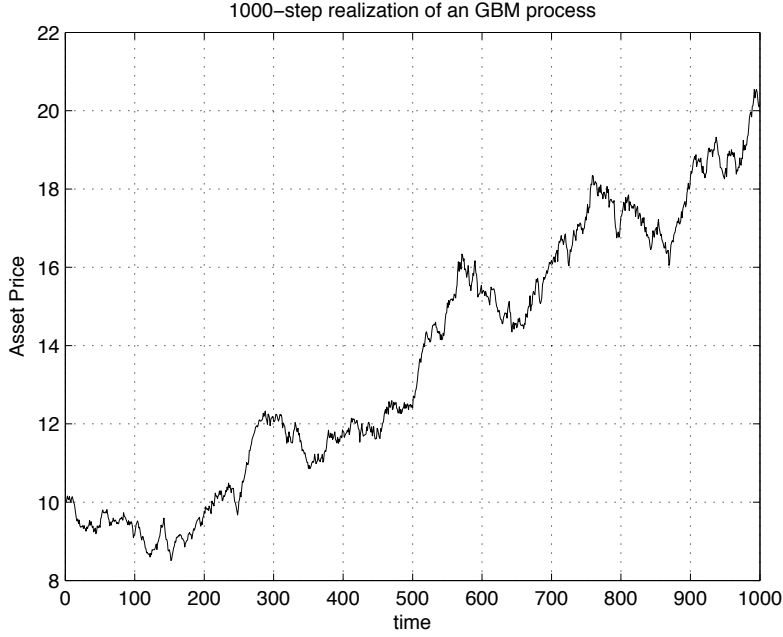


FIG. 4. A realization of a GBM process with $x_0 = 10$, $\mu = 0.25$ and $\sigma = 0.3$.

A critical assumption of GBM is that the price dynamics have continuous sample paths. In order to modify this assumption Merton [35] proposed the so called jump-diffusion process. Under this process the stock price path is discontinuous. The jump diffusion model accepts random price changes of two types. Normal shocks in price are caused by fluctuations in supply and demand and other changes that cause marginal changes in stock price. This is captured by the GBM with constant variance and continuous sample paths. Conversely, abnormal shocks are due to the arrival of important information about the asset that have more than a marginal effect on the price. The times at which such arrivals occur are random. Merton chose a Poisson process as a natural way to represent this jump process, which can be specified with the following SDE

$$(12) \quad \frac{dx_t}{x_{t-}} = \mu dt + \sigma dW_t + dJ_t.$$

The difference with (5) is J , a process independent of W . J is given by

$$(13) \quad J_t = \sum_{j=1}^{N(t)} (Y_j - 1),$$

where Y_1, Y_2, \dots are random variables and $N(t)$ is a counting process. This means that there are random arrival times $0 < \tau_1, \tau_2, \dots$ and

$$(14) \quad N(t) = \sup\{n : \tau_n \leq t\}.$$

In the presence of jumps one must be careful and should differentiate that the process is continuous from the right

$$(15) \quad x_t = \lim_{s \downarrow t} x_s$$

while the value just before the a potential jump is

$$(16) \quad x_{t-} = \lim_{s \uparrow t} x_s.$$

The jump in x at time t is defined by the difference between (15) and (16).

If we consider that x_t evolves as a GBM between jumps, and gets multiplied by J_i at the jump-times $0 < \tau_1, \tau_2, \dots$ of the Poisson process, explicitly

$$(17) x_t = \begin{cases} x_0 \cdot \exp\{(\mu - \frac{1}{2}\sigma^2)dt + \sigma W_t\}, & 0 \leq t \leq \tau_1, \\ x_0 \cdot J_1 \cdot \exp\{(\mu - \frac{1}{2}\sigma^2)dt + \sigma W_t\}, & \tau_1 \leq t \leq \tau_2, \\ x_0 \cdot J_1 \cdot J_2 \cdot \exp\{(\mu - \frac{1}{2}\sigma^2)dt + \sigma W_t\}, & \tau_2 \leq t \leq \tau_3, \\ \text{etc.} & \end{cases}$$

Equivalently, they can be summarized by

$$(18) \quad x_t = x_0 \left(\prod_{k=0}^{N_t} J_k \right) \exp\{(\mu - \frac{1}{2}\sigma^2)dt + \sigma W_t\},$$

where $J_0 = 1$.

In figure 5 we can see a 1000 step realization of a jump-diffusion process with two jumps.

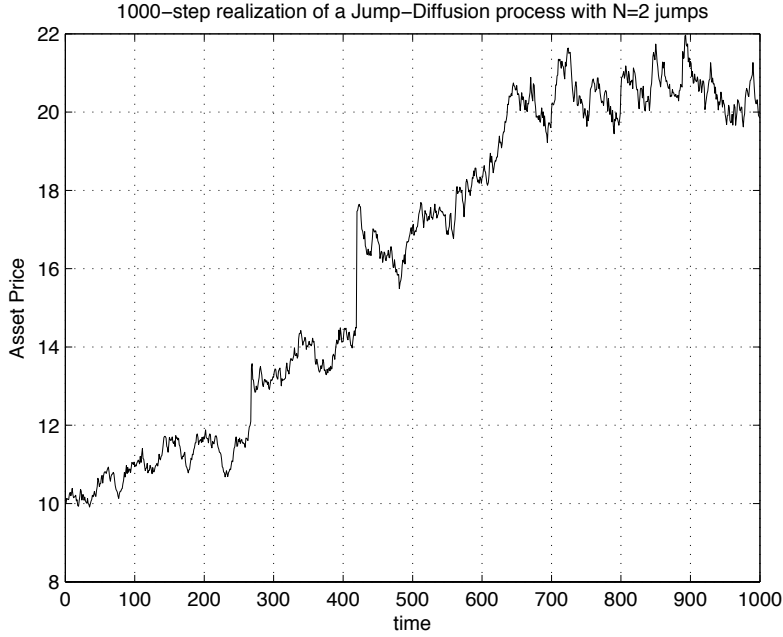


FIG. 5. A realization of a Jump-Diffusion process with $x_0 = 10$, $\mu = 0.25$, $\sigma = 0.3$, $a = 0$, $b = 0.15$ and $\lambda = 3.45$.

5.5. Models for spread dynamics

As seen in the previous section, Brownian motion is suitable to model time series that wander far from their starting points. This is realistic for some economic variables, but not for others. Consider for example the prices of raw commodities such as copper or oil. One should expect them to be somehow related to the long-run marginal production costs. While the short run dynamics might fluctuate randomly, in the longer run it ought to be drawn back to its marginal productions costs. One could apply the same idea, to spreads. Namely, if two economic time series are co-integrated in the long run (i.e: technology, infrastructure, strategy, etc.) one would expect that the difference between them follows two mean reverting process.

Given two time series, say stock A and stock B, quoted in some organized market we compute the spread,

$$(19) \quad x_t = \log(P_t^A) - \log(P_t^B).$$

The idea is to model the increments of the log-difference with the simplest mean-reverting process known as an Ornstein-Uhlenbeck process

$$(20) \quad dx_t = \theta(\mu - x_t)dt + \sigma dW_t,$$

where θ is the speed of mean reversion and μ is the mean level of the process, that is, the level to which x_t tends to revert. Note that if the difference of the expected value and the actual value is positive (negative) the expected change in x , dx , is more likely to rise (fall) over the next short interval of time.

The interpretation of dynamics should be the following: a change in the process is due to a deterministic component, which tends to pull the process back towards the equilibrium asymptote μ at an exponential rate plus a random component dW_t (magnified by the volatility σ) that perturbs this exponential convergence. This behavior is illustrated in figure 6, here the process fluctuates around zero at a high speed of convergence ($\theta = 2$).

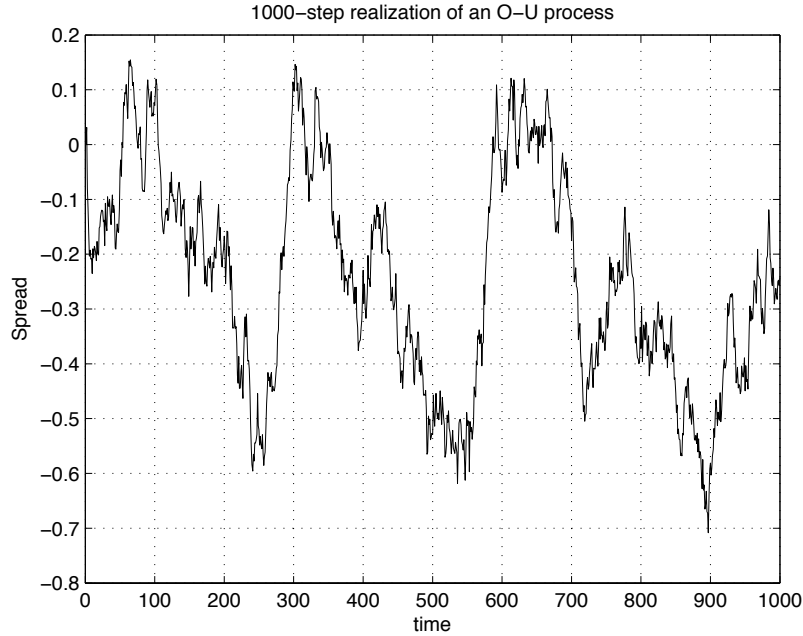


FIG. 6. A realization of an Ornstein-Uhlenbeck process with $\mu = 0$, $\theta = 2$ and $\sigma = 1$.

Now we proceed to derive an analytical solution of the SDE (20). This results will be helpful to build a trading model based on a mean reverting phenomena.

Consider

$$(21) \quad f(x_t, t) = \theta(\mu - x_t) \cdot e^{\theta t}.$$

Using Itô's Lemma

$$(22) \quad df(x_t, t) = \theta(\mu - x_t)e^{\theta t}dt - e^{\theta t}dx_t$$

Replacing (20) into (22) we obtain:

$$\begin{aligned} (23) \quad df(x_t, t) &= \theta(\mu - x_t)e^{\theta t}dt - e^{\theta t}(\theta(\mu - x_t)dt + \sigma dW_t) \\ &= \sigma e^{\theta t}dW_s \end{aligned}$$

Integrating (23) over $(0, t)$, we get

$$(24) \quad f(x_t, t) = f_0 + \sigma \int_0^t e^{\theta s}dW_s$$

which from (21) the explicit solution of the SDE (20) reads:

$$(25) \quad x_t = \mu(1 - e^{-\theta t}) + e^{-\theta t}x_0 + \sigma \int_0^t e^{\theta(s-t)}dW_s$$

The integrand in the Itô integral of equation (25) is nonrandom, so that the process has independent increments and a normal distribution

$$(26) \quad \int_0^t \sigma e^{\theta(s-t)}dW_s \sim \mathcal{N}(0, \frac{\sigma^2}{2\theta}(1 - e^{-2\theta t}))$$

Taking expectations of (25) and assuming that x_0 is constant, we get $\mathbb{E}[x_t] = \mu(1 - e^{-\theta t}) + e^{-\theta t}x_0$

Using the Itô isometry property, stated in (3), we get the variance, $\mathbb{V}[x_t] = \frac{\sigma^2}{2\theta}(1 - e^{-2\theta t})$.

6. Empirical analysis of financial time series

The original form of market prices is tick-by-tick data. Each “tick” is one logical unit of information and by nature these data are irregularly spaced in time. Thus, high-frequency finance should be the primary object of research for those who are willing to understand financial markets.

The idea of this section is to “explore” a data set in order to discover the statistical properties of financial time series and to be able to contrast possible findings with the available models in the literature. Here we will follow a three-step methodology very similar and successful in experimental sciences.

First, the natural step, is to explore the data in order to discover the fundamental statistical properties they exhibit, keeping the assumptions to a minimum. Here the findings are called *stylized facts*, that is, statistical findings of general nature on financial or economic time

series, but not well accepted theories.⁵ In this stage the objective is to formulate adequate models. By adequate models, we mean models that are directly inspired by the empirical regularities encountered in the data. It is here where market behavior and data properties should meet. This step is a key point to validate the model and to satisfactorily test the reproduction of the stylized facts found in the data. Needless to say that a model is judged upon its descriptive capacity and ability to predict the future movements.

The empirical analysis is carried out in the following way. First, we are going to review the statistical properties of financial returns, and second we investigate the logarithmic difference between two time series. The data set is the Dow Jones Industrial Average (DJIA).⁶ The prices are sampled minutely from 01-Jun-2010 to 08-Nov-2010. As the New York Stock Exchange (NYSE) trades from 9:00 to 15:30 (GMT-5) we should have theoretically 44,070 data points. Although the NYSE is very liquid, some equities are not traded with our requested sampling frequency, (i.e: AA and XOM reports 40,711 (lowest) and 44,033 (highest) data points respectively⁷). The missing values are filled with the latest prices available.

6.1. Stock prices returns

Gathering basic stylized facts on the behavior of financial assets and their returns is an important activity, without such facts, is not possible to design models that can explain the data.

Typically, stylized facts can be grouped under the following headings: distributional issues of returns, scaling properties and autocorrelation of returns
indexempirical analysis!autocorrelation of returns.

6.1.1. *Distributional issues.* The objective is not to propose a new model for a probability distribution function for returns but to examine empirically what type of behavior is observed when returns are measured at different frequencies. Typically one can focus on location-dispersion parameters or try to characterize the distribution but looking at the behavior of the tails. This can be described by using only one parameter, the tail index α . Empirical estimation of the tail index requires large numbers of observations, so the best setting is achieved in

⁵i.e: returns of individual stocks exhibit nearly zero autocorrelation at every lag, or the distribution of stock returns is not normal for time horizons from a few minutes to a few days, etc.

⁶It's name derives from the fact that 2/3 of the 30 companies quoted in the index are industrial.

⁷Alcoa and Exxon Mobile Corp.

high frequencies. If P_t represents the price stock at time t , we define the return over the interval size Δt as

$$(27) \quad r_t = \log(P_{t+\Delta t}) - \log(P_t) \approx \frac{P_{t+\Delta t} - P_t}{P_t}$$

The statistical properties of (27) have been studied in many forms over the past 50 years. Fama [14] and Mandelbrot [32] were early challengers to the assumptions of the normality of returns. More recently Stanley and his associates [16, 34, 18] have found tails of financial data corresponding to power law behavior with exponents exceeding that of the Lévy regime.

Ticker	Time interval	Mean	Variance	Skewness	Kurtosis
XOM	1 min	$3.6 \cdot 10^{-6}$	$4.3 \cdot 10^{-7}$	1.60	167
	10 min	$3.5 \cdot 10^{-5}$	$4 \cdot 10^{-6}$	0.76	20
	30 min	$1.1 \cdot 10^{-4}$	$1.1 \cdot 10^{-5}$	0.35	5.8
	60 min	$2.1 \cdot 10^{-4}$	$2.4 \cdot 10^{-5}$	0.45	3.9
CVX	1 min	$3.2 \cdot 10^{-6}$	$5 \cdot 10^{-7}$	2.40	207
	10 min	$3.2 \cdot 10^{-5}$	$4.8 \cdot 10^{-6}$	0.67	24
	30 min	$9.7 \cdot 10^{-5}$	$1.4 \cdot 10^{-5}$	0.30	6.7
	60 min	$1.9 \cdot 10^{-4}$	$3 \cdot 10^{-5}$	0.44	4.4
JPM	1 min	$6.2 \cdot 10^{-7}$	$9.2 \cdot 10^{-7}$	2.10	117
	10 min	$6.7 \cdot 10^{-6}$	$8.5 \cdot 10^{-6}$	0.43	17
	30 min	$2.2 \cdot 10^{-5}$	$2.5 \cdot 10^{-5}$	0.35	7
	60 min	$4.5 \cdot 10^{-5}$	$5.6 \cdot 10^{-5}$	-0.37	8
IBM	1 min	$3.6 \cdot 10^{-6}$	$4.5 \cdot 10^{-7}$	-15	1197
	10 min	$3.6 \cdot 10^{-5}$	$3.9 \cdot 10^{-6}$	-5.4	137
	30 min	$1.1 \cdot 10^{-4}$	$1.1 \cdot 10^{-5}$	-3.1	46
	60 min	$2.1 \cdot 10^{-4}$	$2.4 \cdot 10^{-5}$	-2	24

TABLE 1. Moments of logarithmic return distribution for the most traded securities in the Dow Jones index in the period 01-Jun-2010 to 08-Nov-2010.

Table 1 gives an empirical estimation of the first four moments for the most traded stocks in the sample period, Exxon Mobil Corp., Chevron Corp., JP Morgan and IBM. Sample means and variance are close to zero, but they grow by one order of magnitude when one increases the sample frequency. All the financial time series present the same general behavior. The sample (excess) kurtosis it seems to be a decreasing function of the sample frequency. With the exception of IBM the third moment is below one for time frequencies greater than a minute. We may conclude that the empirical distributions are almost symmetric, the (excess) kurtosis exceeds the value 0, which is the theoretical value

for a Gaussian distribution, and for the shortest time interval, it is considerably high, suggesting that in the large-sample limit may not be finite.

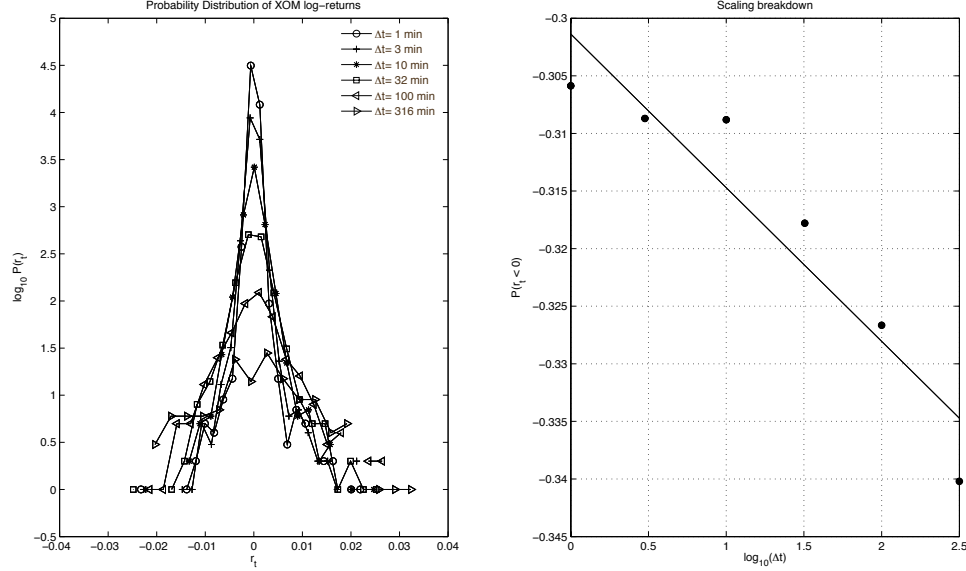


FIG. 7. Probability distributions of XOM log-returns (left). Probability of negative return measured as a function of the time interval Δt (right). No scaling behavior is detected.

Figure 7 (left) reports the empirical probability distribution function (pdf) for the log-returns of XOM for various time interval Δt . In the right panel it can be observed that a power-law behavior previously reported in [34] (on the S&P 500) does not hold for our case of an individual time series.

The tails

indexempirical analysis!tail distribution of all possible distributions can be classified into three categories:

- (1) *Fat-tailed* distributions whose cumulative distribution function declines with a power in the tails.
- (2) *Thin-tailed* distributions for which all moments are finite and whose cumulative distribution function declines exponentially in the tails.
- (3) Bounded distributions which have *no tails*.

The former categories can be distinguished with the use of only one parameter, tail index α , with $\alpha = \infty$ for category (1), $\alpha > 0$ for category (2), and $\alpha < 0$ for category (3).

The rest of the section follows from the framework outlined in [8].

Let X_1, X_2, \dots, X_n be a sequence of n observation drawn from a stationary i.i.d. process whose probability distribution function F is unknown. We assume that the distribution is fat-tailed; that is, the tail index is positive. Let us define $X_{(1)} \geq X_{(2)} \geq \dots \geq X_{(n)}$ as the descending statistics from X_1, X_2, \dots, X_n .

Extreme value theory states that the extreme value distribution of the ordered data must belong to one of the just three mentioned possible general families, regardless of the original distribution function F . Besides, if the original distribution is fat-tailed, there is only one general family it can belong to:

$$(28) \quad G(x) = \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}) & x > 0, \quad \alpha > 0 \end{cases}$$

Where $G(x)$ is the probability that $X_{(1)}$ exceeds X . There is only one parameter to estimate, α .

For doing so, we use the estimator first proposed by Hill [23]

$$(29) \quad \hat{\gamma}_{n,m}^H = \frac{1}{m-1} \sum_{i=1}^{m-1} \ln X_{(i)} - \ln X_{(m)} \quad \text{where } m > 1$$

The estimator was proven to be consistent estimator of $\gamma = \frac{1}{\alpha}$ for fat-tailed distributions.

Ticker	1 min	10 min	30 min	60 min
XOM	2.03	2.76	1.94	1.22
CVX	1.97	3.07	2.11	1.56
JPM	2.15	3.16	2.20	1.41
IBM	2.14	2.78	2.02	1.27

TABLE 2. Estimated tail indices for four Dow Jones stocks with $m=150$.

The results outlined in table 2 are consistent with figure 7 (right panel). The size of the tail seems to be a function of the time interval Δt .

6.1.2. **Scaling properties.** The scaling law is empirically found for a wide range of financial data and time intervals. It gives a direct relation between time intervals Δt and the average volatility measured as a certain power, p , of the absolute returns observed over these intervals,

$$(30) \quad \{\mathbb{E}[|r|^p]\}^{1/p} = c(p) \cdot \Delta t^{D(p)}$$

where \mathbb{E} is the expectation operator $c(p)$ and $D(p)$ are deterministic functions of p . The drift exponent, D , should be equal to 0.5 in order to get a Gaussian random walk, whatever the choice of p .

Taking the logarithm of Equation (30) the estimation of c and D can be carried out by a standard OLS. The results presented here are computed for the cases $p=1$ and $p=2$.

Ticker	$\mathbb{E}[r]$	$\{\mathbb{E}[r ^2]\}^{1/2}$
XOM	0.510	0.479
CVX	0.524	0.492
JPM	0.503	0.484
IBM	0.505	0.470

TABLE 3. Drifts exponents for U.S securities.

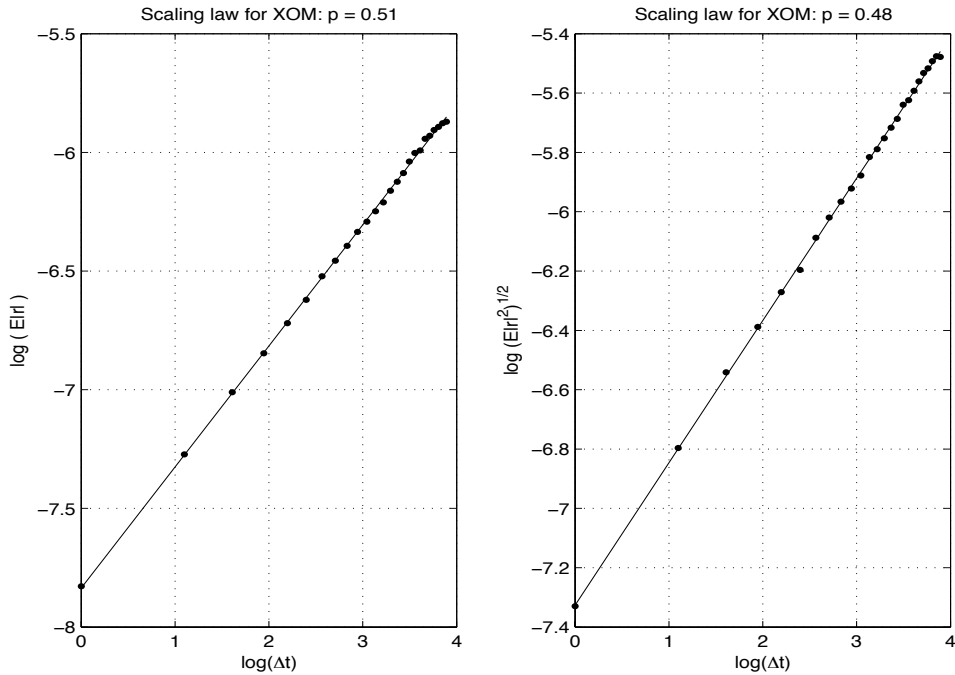


FIG. 8. Scaling law for XOM. The time interval goes from 1 min to 60 min.

6.1.3. ***Autocorrelation of returns.*** Another import issue to classify the behavior of stock returns is the analysis of time correlations. From the results shown in figure 9 we can claim the absence to any first order correlation structure as first proposed in [14].

Figure 10 shows as an interesting picture of the intraday volatility. A well defined pattern emerges indicating a second order time correlation structure. Volatility starts at a medium level at the opening sessions and drops offs after 10:00 and in between 12:00 and 13:00 hits its minimum during lunchtime for most traders. Immediately after it only rises till the closing sessions.

With this pattern in mind we study the second order correlation for XOM, shown in figure 11. Clearly, the volatility exhibit a very strong correlation on the squared returns of the previous closing day and the beginning of the following.

A process with *long memory* is defined as having autocorrelations function,

$$(31) \quad \rho(k) \sim c_\rho |k|^{-\beta},$$

with c_ρ as a constant and k as the lag. A process with this correlation structure indicates that the dependence between far apart events diminishes very slowly with increasing lag. A process can be tested for such a correlation structure by examining the variance of the process' sample mean.

From figure 12 we are able to estimate (least-squares) the long memory parameter β . For the log-return process we find the value $\beta \sim 1$, indicating a lack of memory in the process. However the β value for the squared log-returns provides a strong indication of the presence of long memory in the theses processes.

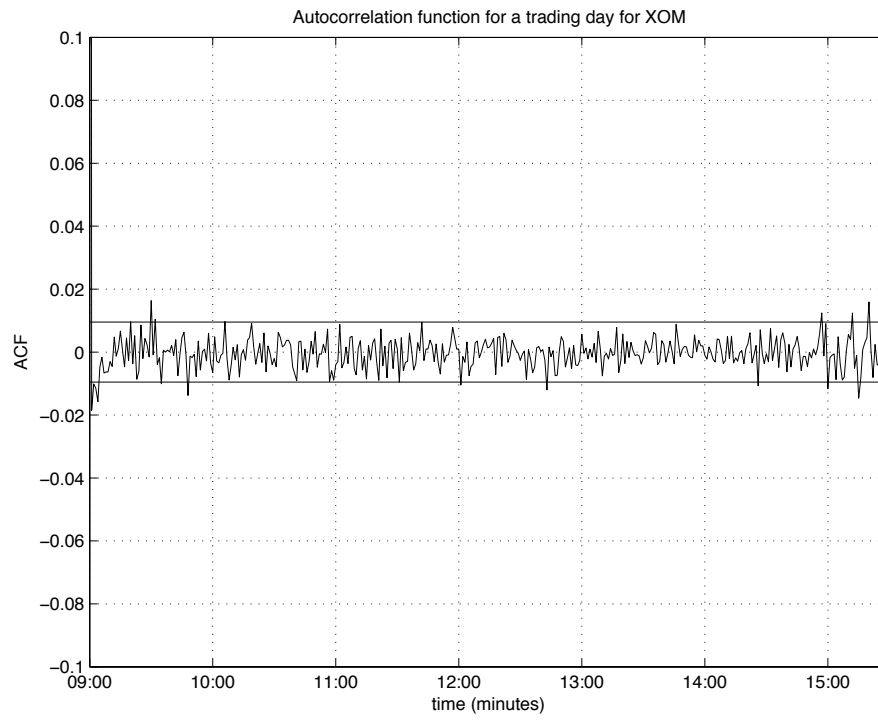


FIG. 9. Intraday autocorrelation function for XOM.

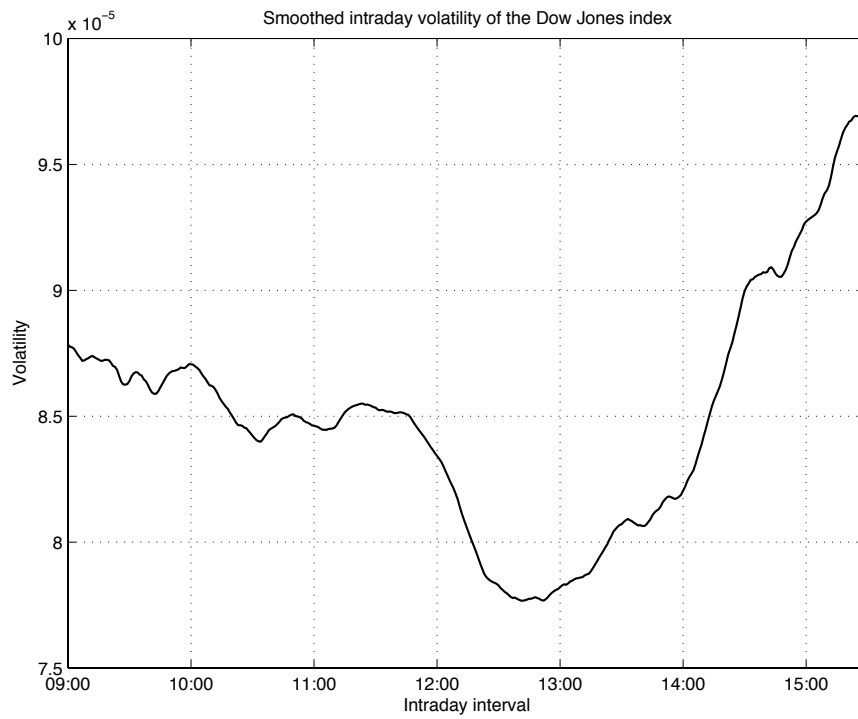


FIG. 10. Intraday volatility profile for the Dow Jones Index.

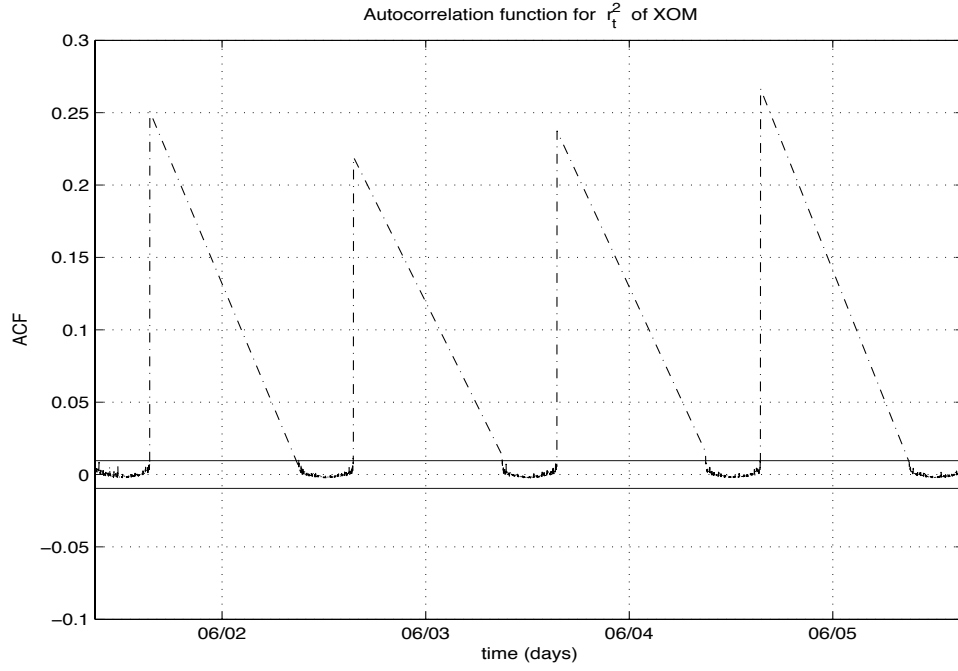


FIG. 11. Intraday autocorrelation function for r_t^2 XOM during a trading week.

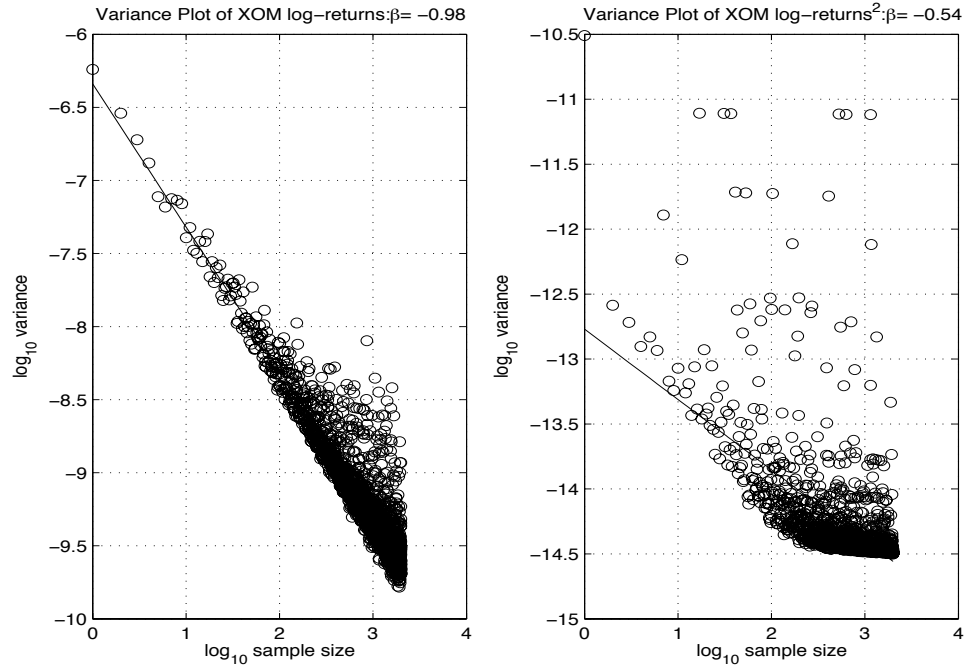


FIG. 12. Variance plot for XOM gives the estimate for the long memory parameter β .

6.2. Spreads between financial assets

The first objective of this section is to study the statistical properties of

$$(32) \quad x_t^{i,j} = \log(P_t^i) - \log(P_t^j),$$

where $x_t^{i,j}$ represent the spread between assets i and j at time t . The second aim is to build an economic taxonomy to select the pairs.

Following Mantegna [33] the starting point to quantify the degree of similarity between the synchronous time evolution of a pair of stock prices would be the correlation coefficient,

$$(33) \quad \rho_{i,j} = \frac{\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle}{\sqrt{(\langle r_i^2 \rangle - \langle r_i \rangle^2)(\langle r_j^2 \rangle - \langle r_j \rangle^2)}},$$

where $\langle \cdot \rangle$ denotes the time average over the investigated period and r_i is given by equation (27). The correlation coefficient $\rho_{i,j}$ can assume values ranging from -1 to 1. If we consider M assets, the correlation matrix contains $\frac{M(M-1)}{2}$ entries, which must be determined from M series of length N . If N is not very large compared to M , one should expect that the determination of the matrix is noisy.

Time period	Minimum	Maximum
2006	0.02	0.83
2007	0.02	0.91
2008	0.02	0.93
2009	0.02	0.87
2010	0.02	0.87

TABLE 4. The observed maximum and minimum values of $\rho_{i,j}$ for the DJIA portfolio for each of the five calendar years from 2006 to 2010.

In order to study the co-movements of stock's returns we investigate the DJIA for a period of the last five years. In table 4 only the minimum and maximum of $\rho_{i,j}$ is reported for each of the time interval. The degree of anti-correlation is practically zero and all maximum values reported are above 0.80. Also table 4 suggest that the density of the correlation coefficient is time-dependent.

In figure 13 we appreciate that the center of the pdf is slowly moving to the right up to 2008, after the financial crisis the pdf widens and

in the current years it seems slowly to recover the former mentioned behavior.

So far, we have discussed the degree of synchronization of the components of the DJIA. Now we want to discuss the relative distance between them. Although the correlation coefficient does not fulfill the requirements to define a metric it helps us to construct one.

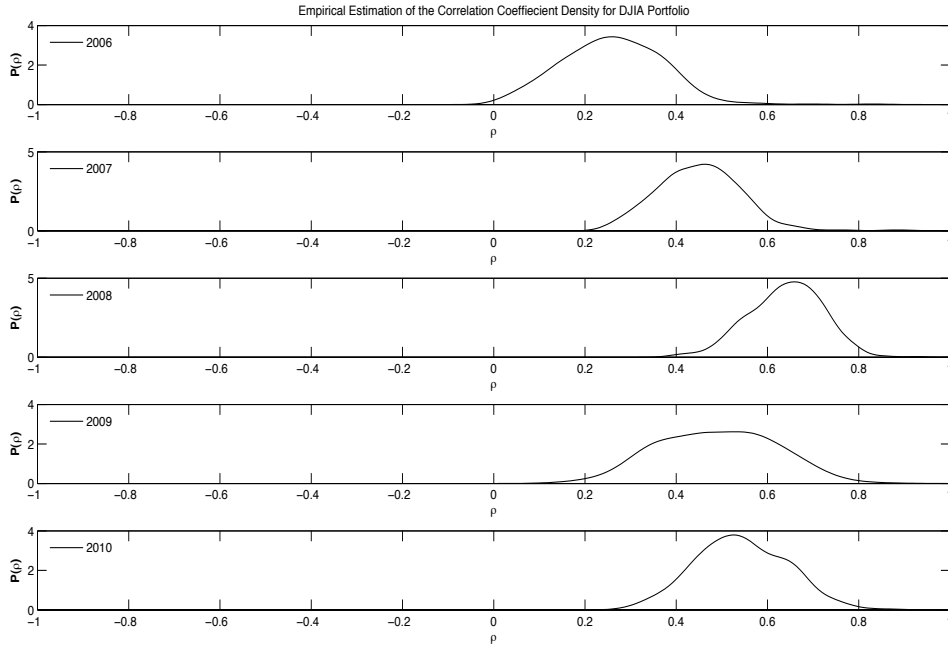


FIG. 13. Probability density function for the correlation coefficients of the DJIA: $P(\rho)$ for each of the five calendar years from 2006 to 2010.

Let us consider the following transformation,

$$(34) \quad \tilde{r}_i = \frac{r_i - \langle r_i \rangle}{\sqrt{N \cdot (\langle r_i^2 \rangle - \langle r_i \rangle^2)}}.$$

Considering the N observations of asset \tilde{r}_i we got an N -dimensional vector $\tilde{\mathbf{r}}_i$. The euclidean distance $d_{i,j}$ between two vectors $\tilde{\mathbf{r}}_i$ and $\tilde{\mathbf{r}}_j$ is obtainable by,

$$(35) \quad d_{i,j}^2 = \|\tilde{\mathbf{r}}_i - \tilde{\mathbf{r}}_j\|^2 = \sum_{k=1}^N (\tilde{r}_{i,k} - \tilde{r}_{j,k})^2.$$

from (34), we can say that $\tilde{\mathbf{r}}_i$ has unit length and

$$(36) \quad \sum_{k=1}^N \tilde{r}_{i,k}^2 = 1.$$

Hence, we can rewrite equation (35) as

$$(37) \quad d_{i,j}^2 = \sum_{k=1}^N (\tilde{r}_{i,k}^2 + \tilde{r}_{j,k}^2 - 2\tilde{r}_{i,k}\tilde{r}_{j,k}) = 2 - 2 \sum_{k=1}^N \tilde{r}_{i,k}\tilde{r}_{j,k}.$$

The sum on the right side of (37), $\sum_{k=1}^N \tilde{r}_{i,k}\tilde{r}_{j,k}$ coincide with $\rho_{i,j}$. Hence the distance between asset i and asset j leads to

$$(38) \quad d_{i,j} = \sqrt{2(1 - \rho_{i,j})}$$

with property $2 \geq d_{i,j} \geq 0$. Because equation (38) defines an Euclidean distance, the following three properties must hold

- (1) $d_{i,j} = 0 \iff i = j$
- (2) $d_{i,j} = d_{j,i}$
- (3) $d_{i,j} \leq d_{i,k} + d_{k,j}$

Properties (1) and (2) are verified because $\rho_{i,j} = 1$ implies $d_{i,j} = 0$, while $\rho_{i,j} = \rho_{j,i}$ implies $d_{i,j} = d_{j,i}$. The triangular inequality relies on the equivalence of equation (37) and equation (38). Thus the quantity $d_{i,j}$ fulfills all three properties that must be satisfied by a metric distance. As stated in [33] one extra hypothesis is necessary to create a useful taxonomy: we need to construct an ultrametric space replacing requirement (3) with a stronger inequality, $\hat{d}_{i,j} = \max\{\hat{d}_{i,k}, \hat{d}_{k,j}\}$.

Ultrametric spaces provide a natural way to describe hierarchically structured complex systems, since the concept of ultrametricity is directly connected with the concept of hierarchy. Among all the possible ultrametric structures associated with a distance metric, $d_{i,j}$, a single one emerges. This is the subdominant ultrametric. In a presence of a metric space with n objects, the subdominant ultrametric can be obtained by determining the minimal-spanning tree (MST) connecting n objects. When constructing the MST, we are effectively reducing the information space from $(M(M-1))/2$ separate distances to M-1 tree edges.

In figure 14 natural groups of the index arise. As we show in figure 13, the time evolution of $\rho_{i,j}$ can be characterized by slow dynamics over a time scale of years. So, one expects that distance is time-dependent, following the same dynamic as $\rho_{i,j}$. We have seen that it is possible to retrieve economic information stored in financial prices and one is able to construct a meaningful taxonomy starting from the synchronous analysis of more than one stock-price time series.

Interestingly, figure 15 gives us a hint about how we can classify the companies pertaining to the DJIA without resorting to any economical

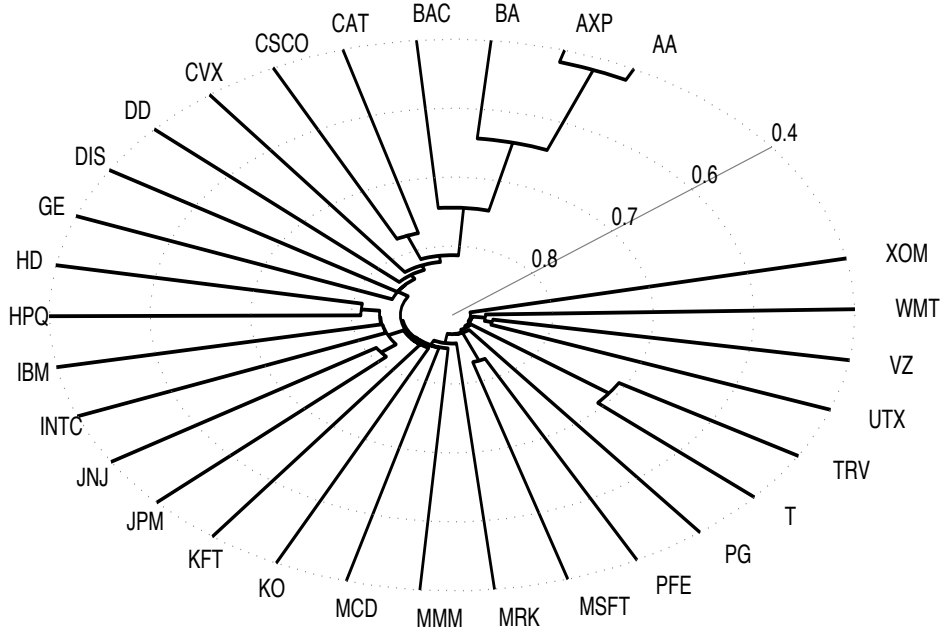


FIG. 14. Taxonomy of the DJIA portfolio. Polar Hierarchical tree.

or financial statements analysis. The idea here, is to explore whether a branch in the tree can be useful to construct schemes to be exploited for trading. Consider for example: DD-AA-CVX-XOM⁸. For another example we can appreciate the information technology cluster formed by HPQ-IBM-INTC-MSFT-CSCO in the upper-left leaf of the minimum spanning tree. .

As we mention before the empirical estimation of the correlation matrix is tricky and could lead us to noisy estimations. Because our distance measure is a function of the correlation coefficient it is important to discuss its estimation. One approach comes from Random Matrix Theory (RMT). An exposition of RMT and financial markets can be found in Bouchaud et al.[4]. Basically, the authors compare the properties of an empirical investigation of $\rho_{i,j}$ to a “null hypothesis”: a purely random matrix as one could obtain from a finite time series of strictly uncorrelated assets.

Deviations from the random matrix case might suggest the presence of true information. Figure 16 illustrates the distribution of the eigenvalues of a random matrix (left) and its limiting distribution (right). The idea is to compare the distribution of the eigenvalues of this matrix with the distribution of the eigenvalues of a correlation matrix.

⁸Du Pont-Alcoa Inc.-Chrevon-Exxon Mobil a cluster formed by materials and oil and gas companies.

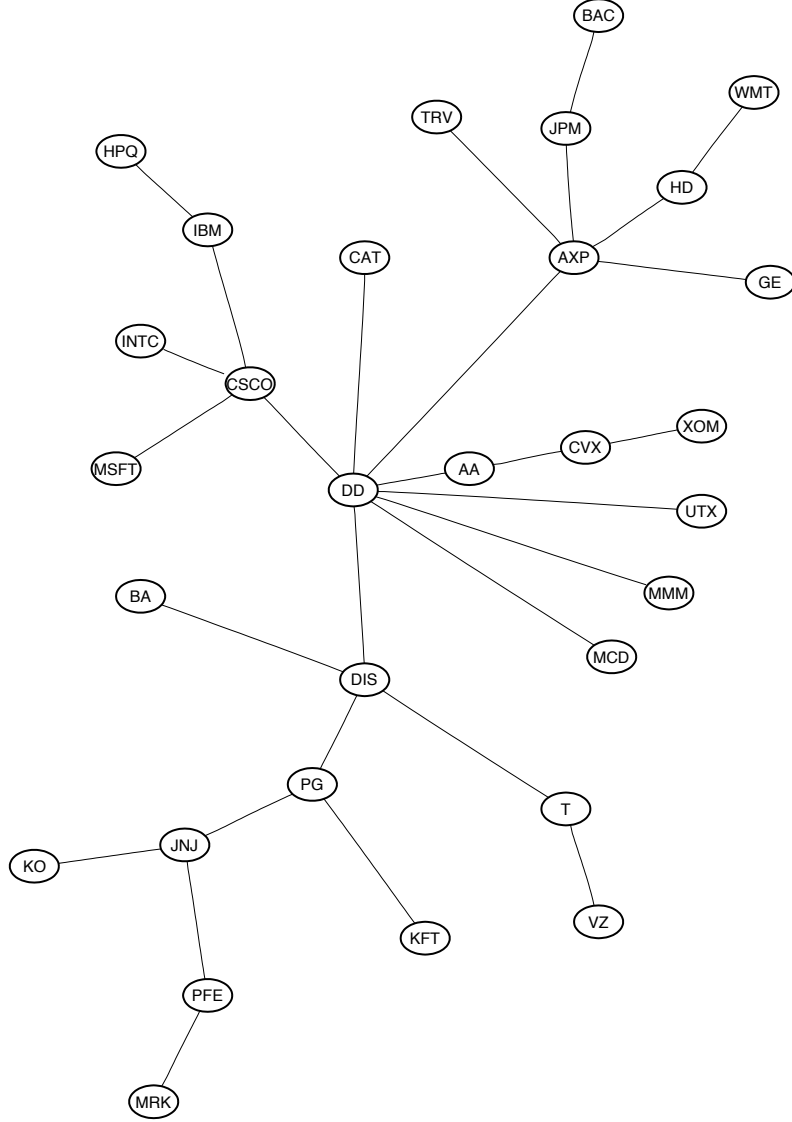


FIG. 15. Taxonomy of the DJIA portfolio. Minimum spanning tree obtained using the Kruskal algorithm.

First, consider the empirical correlation matrix

$$(39) \quad \mathbf{C}_{i,j} = \sum_{k=1}^N \tilde{r}_{i,k} \tilde{r}_{j,k}.$$

Second, we estimate the eigenvalues of matrix \mathbf{C} . The idea is to compare the empirical distribution of eigenvalues of the correlation matrix

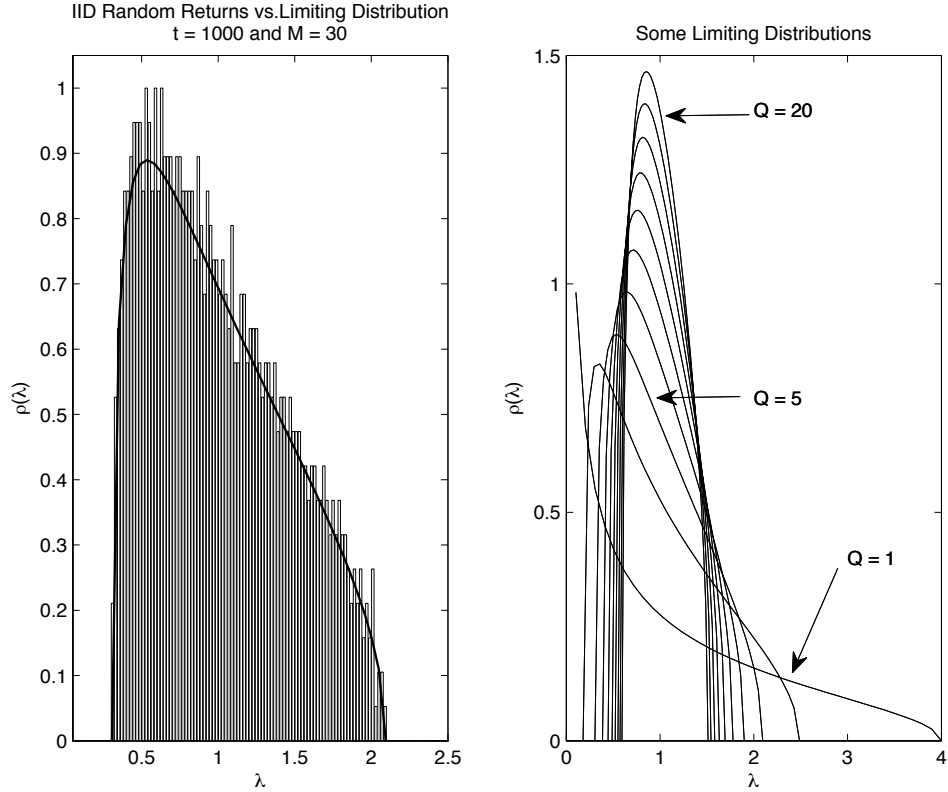


FIG. 16. Empirical distribution of the eigenvalues of a pure random matrix (left). Limiting distributions of the eigenvalues for a random matrix (right).

with the theoretical prediction given by

$$(40) \quad \rho(\lambda_C) = \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_{\max} - \lambda_C)(\lambda_C - \lambda_{\min})}}{\lambda_C},$$

$$\lambda_{\min}^{\max} = \sigma^2(1 + 1/Q \pm 2\sqrt{1/Q}),$$

where $Q = N/M \geq 1$, with $\lambda \in [\lambda_{\min} - \lambda_{\max}]$ and σ^2/N is the variance of the elements of the random matrix.

An immediate observation is that the highest eigenvalue λ_1 is ~ 9 times bigger than the predicted in (40), see figure 17. The corresponding eigenvector is the market itself, i.e: it has equal components on all the N stocks. As the theoretical density does not fit the eigenvalue spectrum of the sample correlation matrix, we may conclude that there is nonrandom structure in the return data as half of the eigenvalues are outside the noise band.

Bouchaud et al. [4] suggests to replace the noise part of the empirical correlation matrix by the identity matrix with a coefficient such that

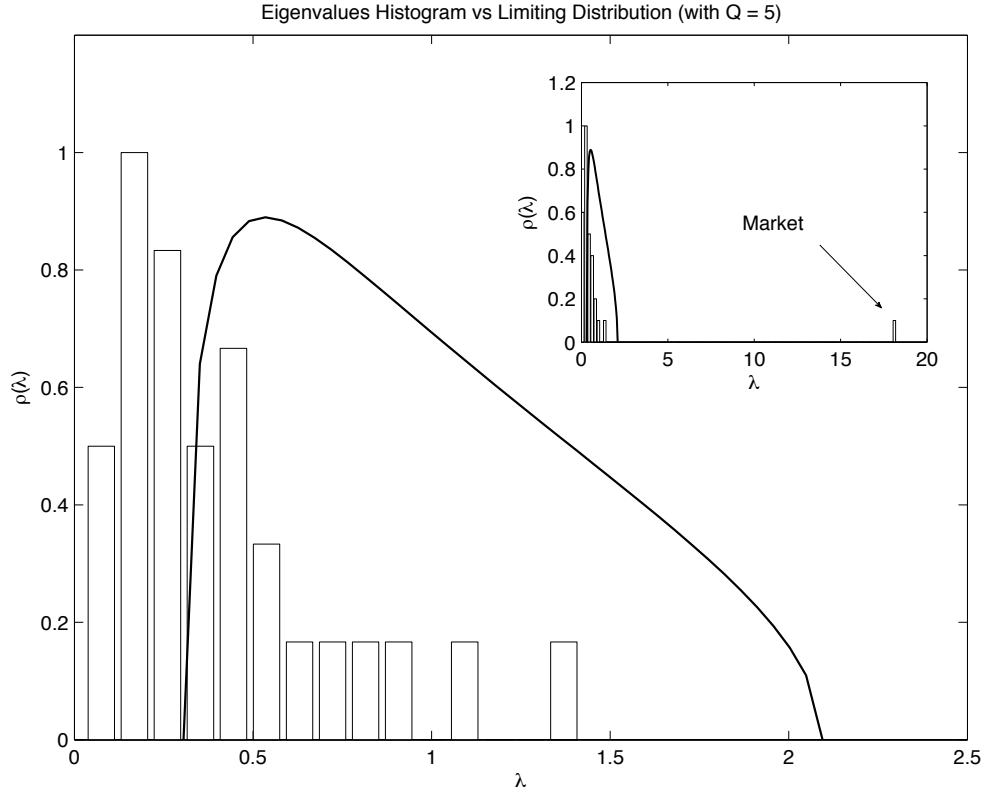


FIG. 17. Histogram of eigenvalues of \mathbf{C} , where the correlation matrix is extracted from the DJIA and the theoretical density for $Q=5$. Inset: same plot, but including the highest eigenvalue corresponding to the ‘market’, which is found to be ~ 9 times higher than λ_{\max} .

the trace of the matrix is preserved. The result is a matrix where the noise has been (partially) removed.

7. Filtering in finance

Nearly all financial models need the specification of parameters. That is, one relies on live market data in order to calibrate and adjust the coefficients of the models to a given situation. However the information needed in the time series of interest may not be fully observable (i.e.: returns, volatilities, risk-premiums, etc.).

The idea behind filtering is to obtain the best possible estimation of a hidden state taking advantage of the continuous update of available information. The concept of filtering has long been used in Control Engineering and Signal Processing. Filtering is an iterative process that enables us to estimate a model’s parameters when the latter relies upon a large quantity of observable and non-observable data.

The idea is to proceed in two steps: first estimate the hidden state *a priori* by using all the information up to that time-step. For then, using this predicted value, together with the new observation, to obtain a conditional *a posteriori* estimation of the state.

The probability density function corresponding to state x_k at time step k , given all the observations $z_{1:k}$ up to that time, is given by applying the Chapman-Kolmogorov equation,

$$\begin{aligned} p(x_k|z_{1:k-1}) &= \int p(x_k|x_{k-1}, z_{1:k-1})p(x_{k-1}|z_{1:k-1})dx_{k-1}. \\ (41) \qquad &= \int p(x_k|x_{k-1})p(x_{k-1}|z_{1:k-1})dx_{k-1}. \end{aligned}$$

Following this, for the *Measurement Update* we use the Bayes rule,

$$(42) \qquad p(x_k|z_{1:k}) = \frac{p(z_k|x_k)p(x_k|z_{1:k-1})}{p(z_k|z_{1:k-1})}p(x_k|z_{1:k-1})$$

Depending whether the model is linear or nonlinear with Gaussian noises, the Kalman Filter (KF) or the Extended Kalman Filter (EKF) it is used respectively. The later is based upon a first order linearization of the transition and measurements equations and therefore coincides with the KF when the equations are linear.

The Kalman Filter is a special case where the distributions are Normal and can be written as,

$$(43) \qquad p(x_k|z_{1:k-1}) \sim \mathcal{N}(\hat{x}_k^-, P_k^-).$$

$$(44) \qquad p(z_k|x_k) \sim \mathcal{N}(\hat{z}_k, P_k).$$

However, it is important to remember that the KF is optimal in the Gaussian *linear case*. In the Non-Linear case, it will be indeed always sub-optimal. When this idea is extended to the Non-Gaussian Case, the Unscented Kalman Filter (UKF) it is used. For a detail exposition of the application of the Kalman filter and its extension, see Javaheri [25].

7.1. Kalman filter

Consider a normally distributed random variable, x , with mean m_x and variance S_{xx} . Also consider another a normally distributed random variable, z , with mean m_z and variance S_{zz} . Having the covariance between x and z , $S_{zx} = S_{xz}$, the conditional distribution of $x|z$ is also normal with

$$(45) \qquad m_{x|z} = m_x + K(z - m_z).$$

Interpreting x as the hidden-state and z as the observable variable, and $K = S_{xz}S_{zz}^{-1}$, as the Kalman gain.

7.2. Extended Kalman filter

For nonlinear Gaussian cases, we can apply the same ideas of the KF via a first-order linearization. The KF could be considered as a particular case of the EKF.

Given a dynamic process \mathbf{x}_k following a non-linear transition equation,

$$(46) \quad \mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{w}_k),$$

we suppose we have a measurement \mathbf{z}_k via a possibly non-linear observation equation,

$$(47) \quad \mathbf{z}_k = \mathbf{h}(\mathbf{x}_{k-1}, \mathbf{u}_k),$$

where \mathbf{w}_k and \mathbf{u}_k are two mutually-uncorrelated sequence of temporally-uncorrelated normal random-variable with zero means and covariance matrices \mathbf{Q}_k and \mathbf{R}_k respectively. We define the *linear a priori* process estimate as:

$$(48) \quad \hat{\mathbf{x}}_k^- = \mathbf{E}[\mathbf{x}_k].$$

which is the estimation at the time step $k - 1$ prior to measurement. Similarly, we define the *linear a posteriori estimate*,

$$(49) \quad \hat{\mathbf{x}}_k = \mathbf{E}[\mathbf{x}_k | \mathbf{z}_k],$$

which is the estimation at time k after the measurement.

We can define the estimation errors $\mathbf{e}_k^- = \mathbf{x}_k - \hat{\mathbf{x}}_k^-$ and $\mathbf{e}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k$ and the estimates error covariances:

$$(50) \quad \mathbf{P}_k^- = \mathbf{E}[\mathbf{e}_k^- \mathbf{e}_k^{-T}],$$

$$(51) \quad \mathbf{P}_k = \mathbf{E}[\mathbf{e}_k \mathbf{e}_k^T],$$

where the superscript T corresponds to the transpose operator.

We now proceed to define the Jacobian matrices of \mathbf{f} with respect to the system process and the system noise as \mathbf{A} and \mathbf{W} respectively. Similarly, we specify the gradient matrices of \mathbf{h} with respect to the system process and the measurement noise as \mathbf{H} and \mathbf{U} respectively.

$$\mathbf{A} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}(\hat{\mathbf{x}}_{k-1}, 0), \quad \mathbf{W} = \frac{\partial \mathbf{f}}{\partial \mathbf{w}}(\hat{\mathbf{x}}_{k-1}, 0),$$

$$\mathbf{H} = \frac{\partial \mathbf{h}}{\partial \mathbf{x}}(\hat{\mathbf{x}}_{k-1}^-, 0), \quad \mathbf{U} = \frac{\partial \mathbf{h}}{\partial \mathbf{u}}(\hat{\mathbf{x}}_{k-1}^-, 0).$$

And the following *Time Update* equations,

$$(52) \quad \hat{\mathbf{x}}_k^- = \mathbf{f}(\hat{\mathbf{x}}_{k-1}^-, 0),$$

$$(53) \quad \mathbf{P}_k^- = \mathbf{A}_k \mathbf{P}_{k-1} \mathbf{A}_k^T + \mathbf{W}_k \mathbf{Q}_{k-1} \mathbf{W}_k^T.$$

Now we define the *Kalman Gain* as the matrix \mathbf{K}_k used in the Measurement equations

$$(54) \quad \hat{\mathbf{x}}_k^- = \hat{\mathbf{x}}_k^- + \hat{\mathbf{K}}_k (\mathbf{z}_k - \mathbf{h}(\hat{\mathbf{x}}_k^-, 0)),$$

$$(55) \quad \mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^-.$$

The optimal Kalman gain corresponds to the mean of the conditional distribution of \mathbf{x}_k upon the observation \mathbf{z}_k , or equivalently, the matrix that would minimize the mean square error \mathbf{P}_k within the class of the linear estimators. The optimal gain is,

$$(56) \quad \mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{U}_k \mathbf{R}_k \mathbf{U}_k^T)^{-1}.$$

7.3. Unscented Kalman filter

A new extension to the KF to nonlinear systems is due to Julier and Uhlmann [27]. They argue that EKF could be difficult to implement and more importantly difficult to tune. The Unscented Kalman Filter (UKF) calculates the mean to a higher order of accuracy than the EKF and the covariance to the same order of accuracy. The underlying idea is that it is easier to approximate a probability distribution than to approximate an arbitrary nonlinear function. It uses an *Unscented Transformation*. By means of constructing a set of Sigma Points (chosen deterministically) which capture the mean and covariance of the original distribution. These points propagated through the nonlinear system, capture the posterior mean and covariance accurately to the third order.

As in the EKF case, we start with an initial choice for the state vector $\hat{\mathbf{x}}_0 = \mathbf{E}[\mathbf{x}_0]$ and its covariance matrix $\mathbf{P}_0 = \mathbf{E}[(\mathbf{x}_0 - \hat{\mathbf{x}}_0)(\mathbf{x}_0 - \hat{\mathbf{x}}_0)^T]$. Then we construct an augmented state vector for each step k greater than one

$$(57) \quad \mathbf{x}_{k-1}^a = \begin{pmatrix} \mathbf{x}_{k-1} \\ \mathbf{w}_{k-1} \\ \mathbf{u}_{k-1} \end{pmatrix}$$

therefore

$$(58) \quad \hat{\mathbf{x}}_{k-1}^a = \begin{pmatrix} \hat{\mathbf{x}}_{k-1} \\ 0 \\ 0 \end{pmatrix}$$

and

$$(59) \quad \mathbf{P}_{k-1}^a = \begin{pmatrix} \mathbf{P}_{k-1} & \mathbf{P}_{xw}(k-1|k-1) & \mathbf{0} \\ \mathbf{P}_{xw}(k-1|k-1) & \mathbf{P}_{ww}(k-1|k-1) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{P}_{uu}(k-1|k-1) \end{pmatrix}$$

for each iteration k . The augmented state will therefore have dimension $n_a = n_x + n_w + n_u$.

Then we need to calculate the corresponding Sigma Points through the *Unscented Transformation*:

$$(60) \quad \begin{aligned} \chi_{k-1}^a(0) &= \hat{\mathbf{x}}_{k-1}^a, \\ \chi_{k-1}^a(i) &= \hat{\mathbf{x}}_{k-1}^a + \left(\sqrt{(n_a + \lambda) \mathbf{P}_{k-1}^a} \right)_i, \quad i = 1, \dots, n_a, \\ \chi_{k-1}^a(i) &= \hat{\mathbf{x}}_{k-1}^a + \left(\sqrt{(n_a + \lambda) \mathbf{P}_{k-1}^a} \right)_{i-n_a}, \quad i = 1, \dots, 2n_a. \end{aligned}$$

The *Time Update* equations are,

$$(61) \quad \chi_{k|k-1}(i) = \mathbf{f}(\chi_{k-1}^x(i), \chi_{k-1}^w(i)), \quad i = 1, \dots, 2n_a + 1,$$

$$(62) \quad \hat{\mathbf{x}}_k^- = \sum_{i=0}^{2n_a} W_i^{(m)} \chi_{k|k-1}(i),$$

and

$$(63) \quad \mathbf{P}_k^- = \sum_{i=0}^{2n_a} W_i^{(c)} (\chi_{k|k}(i) - \hat{\mathbf{x}}_k^-)(\chi_{k|k-1}(i) - \hat{\mathbf{x}}_k^-)^T$$

where the subscripts x and w respectively correspond to the state and system portions of the augmented state. The $W_i^{(m)}$ and $W_i^{(c)}$ weights are defined as

$$(64) \quad W_0^{(m)} = \frac{\lambda}{n_a + \lambda},$$

$$(65) \quad W_0^{(c)} = \frac{\lambda}{n_a + \lambda} + (1 - \alpha^2 + \beta),$$

$$(66) \quad W_i^{(m)} = W_i^{(c)} = \frac{1}{1(n_a + \lambda)}, \quad \text{for } i = 1, \dots, 2n_a.$$

We also define within the *Time Update* equations,

$$(67) \quad \mathbf{Z}_{k|k-1}(i) = \mathbf{h}(\chi_{k|k-1}(i), \chi_{k-1}^u(i)),$$

$$(68) \quad \hat{\mathbf{z}}_k^- = \sum_{i=0}^{(m)} \mathbf{Z}_{k|k-1}(i),$$

where the superscript u corresponds to the observation-noise portion of the augmented state. As for the *Measurement Update* equations, we have

$$(69) \quad \mathbf{P}_{z_k z_k} = \sum_{i=0}^{2n_a} W_i^{(c)} (\mathbf{Z}_{k|k-1}(i) - \hat{\mathbf{x}}_k^-) (\mathbf{Z}_{k|k-1}(i) - \hat{\mathbf{x}}_k^-)^T,$$

$$(70) \quad \mathbf{P}_{x_k z_k} = \sum_{i=0}^{2n_a} W_i^{(c)} (\chi_{k|k-1}(i) - \hat{\mathbf{x}}_k^-) (\mathbf{Z}_{k|k-1}(i) - \hat{\mathbf{x}}_k^-)^T,$$

which gives us the Kalman Gain

$$(71) \quad \mathbf{K}_k = \mathbf{P}_{x_k z_k} \mathbf{P}_{z_k z_k}^{-1}$$

and we get as before

$$(72) \quad \hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k (\mathbf{z}_k - \hat{\mathbf{z}}_k^-),$$

where again z_k is the observation at time (iteration) k .

Also we have

$$(73) \quad \mathbf{P}_k = \mathbf{P}_k^- - \mathbf{K}_k \mathbf{P}_{z_k z_k} \mathbf{K}_k^T,$$

which completes the Measurement Update Equations.

8. Statistical testing of trading strategies

To make the final (irreversible) decision on funding a quantitative trading strategy a statistical test must be carried out in order to differentiate whether the strategy is profitable due to skills or due to chance.

As suggested by Brock et al. [7] it is possible to conduct statistical testing on the trading strategies at two different levels of depth. At a **first stage**, one can compare profits coming from buy and sell trades. If a certain strategy does not have any predicting power, then the profits on buying days should no differ from profits on selling days. Therefore we can use standard statistical tests and compare profits generated by a given strategy and a passive one such as buy-and-hold.

For example, by using the t-statistic we can check if the mean buy and mean sell profits are significantly different. The t-statistic is

$$(74) \quad t^{b,s} = \frac{\mu_b - \mu_s}{(\sigma_b^2/N_b + \sigma_s^2/N_s)^{1/2}},$$

where μ_b (μ_s), σ_b^2 (σ_s^2) and N_b (N_s) are respectively the mean profit, the number of signals and the variance of profits for buys (sells).

In a similar fashion it can be calculated a t-statistic to test whether the mean of certain strategy is significantly different from a buy-and-hold statistic. These tests assume normal, stationary and time-independent distributions. These assumptions are however too strong when compared to the stylized facts which affect stock returns distributions and financial time series. Therefore standard statistical tests can thus give misleading results.

At a **second stage**, a possible solution is to bootstrap data. The idea is to estimate confidence intervals for trading profits under various null models (i.e: the random walk, the AR(1), the GARCH(1,1), etc.) whose parameters will be estimated on the data.

By using the bootstrapping procedure, standardized residuals are not restricted to a particular distribution, and we can test a particular strategy against different models.

Following [7] the procedure is as follows. Once the null model of choice has been estimated on the data set and its residuals standardized, re-sample with replacement and use them as innovations to simulate a new price history of the same length as the original. After that we have to run the trading strategy to the bootstrapped series and repeat n times to obtain the empirical distributions of mean profits of the strategy under the chosen generating model.

We can compute the p-values as fraction of simulated strategies which have greater mean or standard deviation than the strategy applied to the original series.

9. Performance measurements of financial investments

9.1. Measuring the risk of financial investments

If we were to define a measure of overall risk Φ of a position α , it must fulfill the following requirements [1]:

- (1) **Translation invariant.** Adding a fixed amount of wealth (δ) with certainty to the position should decrease the risk associated with that position by the same amount of wealth

$$\Phi(\alpha + \delta) = \Phi(\alpha) - \delta$$

- (2) **Monotonous.** If a position α is always better than another one β , the risk involved in the former should be lower

$$\Phi(\alpha) \leq \Phi(\beta).$$

- (3) **Positively homogeneous.** The risk associated with h times the position should be h times the risk

$$\Phi(h\alpha) = h\Phi(\alpha) \quad \forall h > 0.$$

- (4) **Sub-additive.** The risk of a portfolio should not exceed the sum of the risk of its components

$$\Phi(\alpha + \beta) \leq \Phi(\alpha) + \Phi(\beta)$$

Henceforth, a coherent measure of risk will be meant any functional $\Phi : \mathcal{L}^2(\Omega) \rightarrow (-\infty, \infty)$ satisfying the above axioms.

Examples of non-conforming measures of risk are the Value at Risk (VaR)⁹, which does not respect the sub-additive property. Conversely, the expected shortfall (ES) satisfy all the above requirements and it is indeed a coherent measure of risk. The ES is simply defined as the average among the worst $(1 - \alpha)\%$ cases, formally defined as

$$(75) \quad \text{ES}(\alpha) = \frac{1}{1 - \alpha} \int_0^{1-\alpha} F^{-1}(u) du,$$

where F is the cumulative distribution function (cdf).

The most common measure of risk-adjusted return is the so-called [Sharpe ratio](#) (SR). Basically an indicator of excess return per unit of deviation, in this case the standard deviation;

$$(76) \quad \text{SR} = \frac{\langle r \rangle - r^{\text{free}}}{\sigma}.$$

It is fair to ask if the standard deviation of the returns of a given investment, (σ) or semi-deviations (σ_-, σ_+) ¹⁰, are coherent measures of risk. For a deviation measure we mean any functional $\mathcal{D} : \mathcal{L}^2(\Omega) \rightarrow [0, \infty)$ satisfying

- (1) $\mathcal{D}(X + C) = \mathcal{D}(X)$, $\forall X$ and constant C .
- (2) $\mathcal{D}(0) = 0$, and $\mathcal{D}(\lambda X) = \lambda \mathcal{D}(X)$, $\forall \lambda > 0$
- (3) $\mathcal{D}(X + Y) \leq \mathcal{D}(X) + \mathcal{D}(Y)$
- (4) $\mathcal{D}(X) \geq 0$

⁹ $\text{VaR}(\alpha) \equiv F^{-1}(1 - \alpha) = x$

¹⁰where $\sigma_+ = \|[X - EX]_+\|$ and $\sigma_- = \|[X - EX]_-\|$.

The above properties hold for σ (symmetric) and σ_-, σ_+ (not symmetric). Deviations or semi-deviations are not risk measures in the sense of [1]. Instead of measuring the uncertainty in X , a risk measure evaluates the “overall seriousness of possible losses”¹¹ associated with X . In other words, there is no functional on $\mathcal{L}^2(\Omega)$ that can satisfy both first requirements of a risk and deviation measure.

For an exposition on the relationship between deviations measures and risk measures see [44]. For our own interest it will suffice to clarify that when we talk about performance or “adjust” performance the risk is not taking into account coherently.

When returns are not symmetric, the upside risk is penalized with the same weight as the downside risk. To fix this issue, Sortino introduces a slight modification to the SR. Instead he suggests using a downside deviation measure $\sigma_T = \|X - T\|_p$, where T is the target return (i.e: risk-free return) and $\|\cdot\|_p$ is the norm in \mathcal{L}^p .

Conversely to the previous performance indicators, [Modigliani and Modigliani](#) (MM) [37] suggest a risk-adjusted indicator reported in basis points that compares a certain strategy against a relevant unmanaged “market” portfolio.

$$(77) \quad \text{MM} = \frac{\sigma_M}{\sigma} (\langle r \rangle - r^{\text{free}}) + r^{\text{free}},$$

where σ_M is the standard deviation of the market or benchmark portfolio. The advantage of using MM as a measure of performance is that provides a much more intuitive figure (in basis points) after normalizing the deviation with respect to the market in which the transactions are being executed.

9.2. Wealth Dynamics

Due to Shannon [45], the fundamental theorem of information theory specifies a special rate at which binary digits can be transmitted with an arbitrary small probability of error. When applying this ideas to maximizing the rate of log-growth the wealth of a certain portfolio, the optimum, is found to be Shannon’s special rate of transmission. While working at Bell Labs, Kelly [28] developed this ideas and generalized to situations where one have noisy predictions. The [Kelly criterion](#) (KC) specifies a way of asymptotically optimizing capital growth by means of taking advantage compounding over the long run.

¹¹A loss is an outcome below 0. If the concern is over the extent to which a given r.v X might have outcomes $X(\omega)$ that drop below a threshold C , ones replaces X with $X - C$.

Suppose that V_0 is the value of some initial wealth invested in a risky asset. During each period $i = 1, 2, \dots, n$ our strategy makes iid random returns governed by (5), r_i . The wealth at the end of the period T is

$$\begin{aligned} V_T &= V_0 \prod_{i=1}^T (1 + r_i) \\ (78) \quad &= V_0 \exp \left(\sum_{i=1}^T \log(1 + r_i) \right) \end{aligned}$$

The compound growth rate, CGR, is the average rate of return per period

$$\begin{aligned} \text{CGR}_T &= (V_T/V_0)^{1/n} - 1 \\ &= \exp \left(\sum_{i=1}^T \log(1 + r_i) \right)^{1/n} - 1 \\ &= \exp \left(\frac{1}{n} \sum_{i=1}^T \log(1 + r_i) \right) - 1 \\ (79) \quad \lim_{n \rightarrow \infty} \text{CGR}_n &= \exp \langle \log(1 + r_i) \rangle - 1 \end{aligned}$$

If we are able to split V_t between a risk-free asset (i.e: cash) and a log-normally distributed risky asset, and we continuously rebalance the proportion of the wealth, f , then the portfolio instantaneous rate of changes is governed by

$$\begin{aligned} \frac{dV_t}{V_t} &= (1 - f)r^{\text{free}}dt + f(\mu dt + \sigma dW_t) \\ (80) \quad &= (r^{\text{free}} + f\sigma\lambda)dt + f\sigma dW_t, \end{aligned}$$

where λ is the Sharpe Ratio. Applying Itô's lemma we obtain

$$(81) \quad d\log V_t = \underbrace{(r^{\text{free}} + f\sigma\lambda - \frac{1}{2}f^2\sigma^2)}_{g(f)=\log\text{-growth rate}}dt + f\sigma dW.$$

Maximizing the logarithmic growth rate we get,

$$\begin{aligned} \frac{\partial}{\partial f} g(f)|_{f=f^*} &= \sigma\lambda - f^*\sigma^2, \\ (82) \quad &= 0. \end{aligned}$$

And so, $f^* = \lambda/\sigma$. This suggests the following log-optimal wealth dynamics:

$$(83) \quad \frac{dV_t}{V_t} = (r^{\text{free}} + \lambda^2)dt + \lambda dW_t.$$

The volatility of the log-optimal portfolio is equal to the SR. Intuitively, the optimal quantity f^* may seem overly aggressive. If we consider a strategy with volatility of 20% and unitary SR, the KC prescribes placing five times our initial wealth in the risky asset by shorting the risk free asset (borrowing). This is the reason why some may choose only a fraction of the ratio suggested by the KC. Figure 18 illustrates how an investor can choose a certain *Kelly fraction* in order to maximize the logarithmic growth of his portfolio.

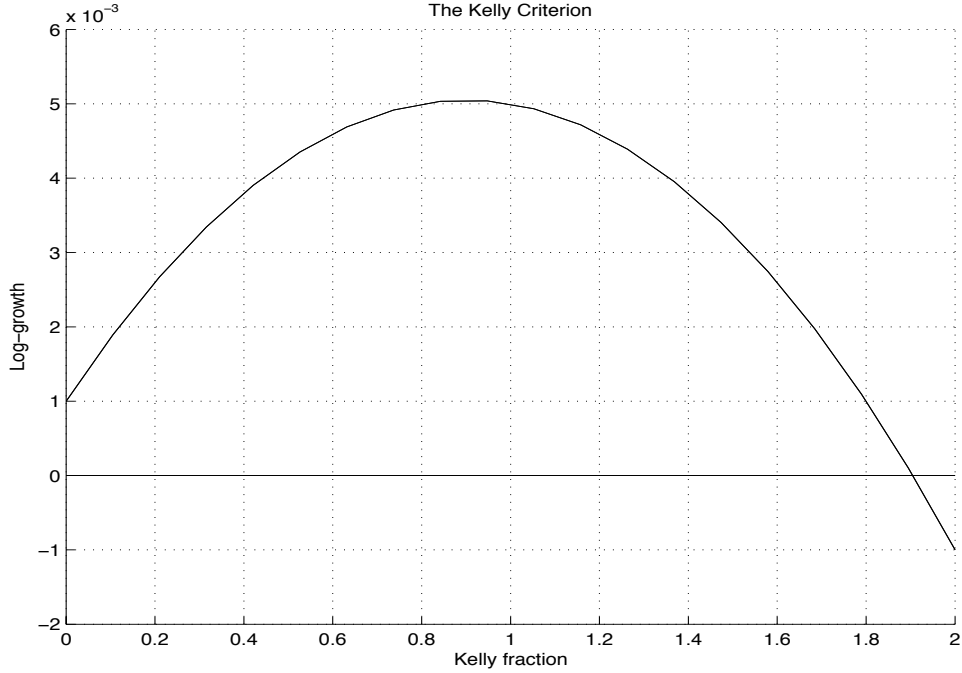


FIG. 18. Full Kelly investor choose $f=1$ which is equivalent to maximize the log-utility function $u(x) = \log(c)$. Partial Kelly strategies ($f < 1$) are equivalent to maximize $u(x; \alpha) = \alpha \cdot x^\alpha$, with $\alpha < 1$

Consider a fractional Kelly investor who maximize a negative power utility, the optimal kelly fraction is $f^* = \frac{1}{1-\alpha}$. The dynamics of the utility function is governed by

$$(84) \quad \frac{du_t}{u_t} = [\alpha(r^{\text{free}} + f^2\sigma^2) + \alpha(\alpha - 1)f^2\sigma^2/2] dt + \alpha\sigma dW_t$$

The expected terminal value of the utility is given by

$$(85) \quad \langle u_T |_{t=0} \rangle = u_0 \exp [T\alpha(r^{\text{free}} + f^2\sigma^2) + T\alpha(\alpha - 1)f^2\sigma^2/2]$$

So the value f that maximize the terminal utility is given by,

$$\begin{aligned}
\frac{\partial}{\partial f} g(f)|_{f=f^*} &= T\alpha\sigma^2 + T\alpha(\alpha-1)f^*\sigma^2, \\
&= 0, \\
(86) \qquad f^* &= \frac{1}{1-\alpha}.
\end{aligned}$$

Chapter 2

Technical Trading

The term *technical analysis* (TA) is a general myriad of trading techniques. Technical analysts attempt to forecast prices by the study of past prices and a few related summary statistics about security trading. Basically, TA rely on the fact that financial markets are not random. Dynamic strategies could exploit this imperfections and outperform the market. Namely, serial dependences could be exploited by directional trading rules. Despite that TA does not enjoy a healthy reputation within the academic community, it is worthy to provide a framework in order to answer the question whether TA can be formalized or if it lack of total scientific value. As mentioned in the previous chapter, the research conducted in [40] tackles this issue.

Let $\{X_t\}_{(t=0,1,\dots)}$ be an asset price quoted in some financial market. Let $\{\mathcal{F}_t\}_{(t=0,1,\dots)}$ be the sigma-algebras generated by the X_t and possibly by other data observed up to time t .

Definition 14. We say that a random variable τ is a Markov or stopping time if the event $A_t = \{\tau < t\}$ is \mathcal{F}_t -measurable.

Example 1. Let τ_1 denote the date at which a process, observed continuously $\{X_t\}$ shows a 5% jump for the first time during $t \in [0, \infty)$:

$$(87) \quad \tau_1 = \inf_t \left\{ t \in [0, \infty) : \frac{d(\ln X_t)}{dt} > 0.05 \right\}.$$

Then τ_1 is a Markov time since, by looking at the current information set, it is possible to tell whether such a jump in X_t has occurred or not.

Example 2. Let τ_2 denote the beginning of a market uptrend. Then, τ_2 is not a Markov time since, in order to know whether $t = \tau_2$, one needs to have access to \mathcal{F}_{t+s} , with $s > 0$.

Clearly, any well-defined [technical analysis](#) rule has to pass the test of being a Markov time since any buying or selling signal should, in principle, be an announcement based on data available at time t . If a rule generates a sequence of buy and sell orders that fails to be Markov

times, then the procedure would be using *future information* in order to issue such signals.

Graphic methods are not the best way of determining the classes of Markov times that are useful in prediction. Yet, more often than not, this is how TA rules are defined. Hence, the importance of developing formal algorithms that can duplicate the buying and selling signals given by technicians.

If one can show that signals generated by a rule of TA are Markov times, then this would imply: (1) that the method can be quantified, (2) that it is feasible, and (3), that one can investigate its predictive power using formal statistical models.

Equipped with a criterion to determine which rules of technical analysis can be quantified, we are now ready to review the most common indicators.

1. Moving average crossings indicators

This rule of TA uses sign changes to generate the Markov times $\{\tau_i\}$.

First, we define:

$$(88) \quad Z_t = \left[\frac{1}{n} \sum_{s=0}^{n-1} X_{t-s} \right] - \left[\frac{1}{m} \sum_{s=0}^{m-1} X_{t-s} \right].$$

The sequence is generated by means of:

$$(89) \quad \tau_i = \inf_t \{t : t > \tau_{i-1}, \quad Z_t Z_{t-1} < 0\}.$$

Basically, in (88), two moving averages (MA) of the X_t process are calculated, assuming that $n > m$. The first moving average will be smoother and carries more information. Then, as soon as Z_t , for $\tau_{i-1} < t$ changes sign, the rule (89) assigns the value of t to τ_i .

The product $Z_t Z_{t-1}$ is measurable with respect to \mathcal{F}_t . That is, given the value of \mathcal{F}_t , the value of $Z_t Z_{t-1}$ is known. The τ_i sequence is then well defined. This makes the moving average crossing method a well defined procedure.

Figure 1 shows us a clear example of how a moving average crossing works: around december 2006 the longer MA was above the shorter MA, which clearly indicated a selling signal.

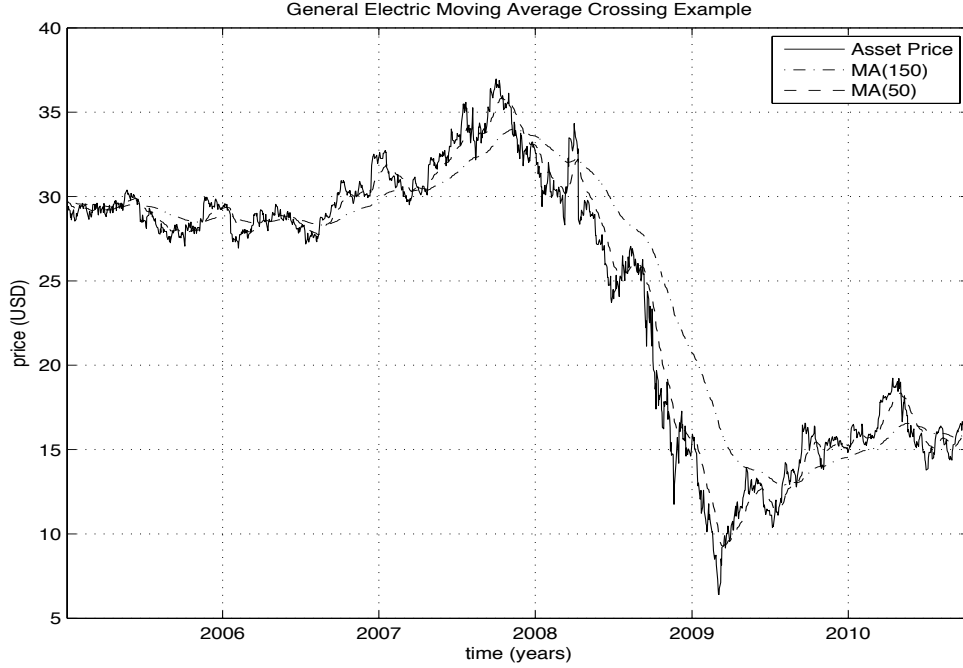


FIG. 19. Moving average crossing for General Electric (upper panel). Trading volume (lower panel).

2. Trend crossing methods

The notion of moving average immediately suggests a mathematical formulation, whereas trend crossings seems to be based on arbitrary hand-drawn trends in charts illustrating historical data. The main idea behind trend crossing methods is to determine two linear trends, one above, the other below, that would envelope the portion of data observed since the last turning point.

Then, up-crossings (down-crossings) of the upper (lower) envelope are taken as signals of market strength (weakness).

These trend lines can be defined by using local extrema of a portion of series under consideration.

Consider the trend line $T(t)$:

$$(90) \quad T(t) = \left[\frac{X_1 - X_0}{t_1 - t_0} \right] (t) + \left[\frac{X_0 t_1 - X_1 t_0}{t_1 - t_0} \right].$$

for $t_1 > t_0 > \tau_{i-1}$. This function defines a straight line that goes through the two lowest (highest) local minima (maxima) observed during the interval $(\tau_{i-1}, t]$. The Markov times are generated using:

$$(91) \quad Z_t = X_t - T(t),$$

and

$$(92) \quad \tau_i = \inf_t \{t : t > \tau_{i-1}, \quad Z_t Z_{t-1} < 0\}.$$

It seems from (91) that Z_t is \mathcal{F}_t -measurable. And if we adopt the rule from (92) to determine the $\{\tau_i\}$, then this would be the first entry in the interval $[0, \infty)$ by a \mathcal{F}_t -measurable random variable. Hence, the τ_i would be Markov times. But it turns out that, in general, Z_t is not a \mathcal{F}_t -measurable since t_0 and t_1 are never specified as the times of the onset of the first two (or the n th) local extrema during $t > \tau_{i1}$. This implies that the trend crossing technique will not generate Markov times unless one specifies a \mathcal{F}_t -measurable mechanism for ignoring the local minima (maxima) between t_0 and t_1 .

3. Technical oscillators

Technical analysts use oscillators to discover short-term overbought or oversold situations. Among the most popular used by practitioners we can find: the “rate of change” (ROC), “the commodity channel index” (CCI), “the stochastic index” (SI) and the “relative strength index” (RSI), to name a few. Here we will study the later. The RSI measures the velocity and magnitude of the directional price movements.

Let δ be an exponential moving average (EMA) given by:

$$(93) \quad \delta_t = \frac{2}{n+1} \cdot (X_t - \delta_{t-1}) + \delta_{t-1}, \quad \text{for } t > 2,$$

where,

$\delta_1 = X_1$, and n is the length of the average.

The next step is to divide δ_t in two sets: $\delta_t^u = \max(P_t - P_{t-1}, 0)$ and $\delta_t^d = \max(X_{t-1} - X_t, 0)$. The first set represent the upward movement of the financial time series. Conversely, the second set represent the downward movements of two consecutive data points. Next we calculate the relative strength ratio, defined as: $RS_t = \delta_t^u / \delta_t^d$. Finally we adjust this indicator to lay between the interval $[0, 100]$.

Then, the RSI reads:

$$(94) \quad RSI_t = 100 - \frac{100}{1 + RS_t}.$$

Now, equipped with the RSI, we need to establish the threshold λ , by which the Markov times or trading signals are activated. Typically, λ is set to be in the interval $[55, 70]$.

A selling signal is activated when the RSI hits the threshold λ and conversely a buying signal is represented when the RSI falls below $1 - \lambda$. The Markov times are generated by:

$$(95) \quad Z_t = (\lambda - \text{RSI}_t) \cdot (100 - \lambda - \text{RSI}_t),$$

and

$$(96) \quad \tau_i = \inf_t \{t : t > \tau_{i-1}, \quad Z_t > 0\}.$$

As this indicator it is based on a moving average method, the sequence generated by Z_t are also valid Markov times and represents a \mathcal{F}_t -measurable sequence.

4. Patterns in financial time series

Another class of procedures used by technical analysts involve various patterns to issue signals. Two of the most popular patterns are considered, namely, “triangles” and “head and shoulders”. Although these patterns, in principle, can be formally defined using particular sequences of local minima and maxima, they are not \mathcal{F}_t -measurable events.

5. Moving average convergence divergence

The moving average convergence divergence (MACD) combines two exponential moving averages of past prices into two lines: the MACD line and the signal line (SL). The MACD line is constructed as the difference between two exponential moving averages computed using the last m and n prices, where m and n are integers such that $n > m$. It follows that the MACD line crosses the zero line each time there is a crossover between the two moving averages. Let

$$(97) \quad \delta_t^n = w_1 \cdot (X_t - \delta_{t-1}^n) + \delta_{t-1}^n, \quad \text{for } t > 2,$$

$$(98) \quad \delta_t^m = w_2 \cdot (X_t - \delta_{t-1}^m) + \delta_{t-1}^m, \quad \text{for } t > 2,$$

be the long-period and short-period exponential moving average, respectively with $\delta_1^n = X_1$, $w_1 = \frac{2}{n+1}$ and $\delta_1^m = X_1$, $w_2 = \frac{2}{m+1}$ and

$$(99) \quad \text{MACD}_t = \delta_t^m - \delta_t^n, \quad \text{MACD}_1 = 0.$$

It is easy to observe that, since the fast moving average (m periods) reflects the short-period trend in the market, while the slow moving average (n periods) reflects the long-period tendency, when the MACD is positive it can be interpreted as a bullish market phase.

We now define the so-called signal line

$$(100) \quad \text{SL}_t = \frac{1}{k} \text{MACD}(t) + \left(1 - \frac{1}{k}\right), \quad \text{SL}_1 = 0.$$

According to the MACD rules, a buying (selling) signal is generated each time the MACD line moves above (below) the SL. Hence the Markov times are generated by

$$(101) \quad \tau_i = \inf_t \{t : t > \tau_{i-1}, \quad Z_t Z_{t-1} < 0\}.$$

where $Z_t = \text{MACD}_t - \text{SL}_t$.

Unlike other oscillators, like the RSI or the SI, MACD is not constrained between upper and lower bounds, but it can hit new highs or lows as long as the trends are gaining momentum (rising asset prices). Since this trading rule relies on moving averages, it generates legal signals in the sense of Markov times.

6. Mixed indicators models

Very often technical analysts combine these indicators to average the signals. Here we will review a combo model formed by a moving average plus an oscillator both of them giving sequences of Markov times.

In this case the Markov times are generated by means of:

$$(102) \quad Z_t = (Z_t^e + Z_t^r)/2,$$

and

$$(103) \quad \tau_i = \inf_t \{t : t > \tau_{i-1}, \quad |Z_t| < 1\},$$

where, $Z_t^e = X_t - \delta_t^n$ and Z_t^r it is defined according to (95).

As mentioned in sections 1 and 3, both moving average and oscillators constitute well defined methods and do not fail to generate Markov times.

7. Finite Markov times

Once it has been identified a technical analysis rule that it is known to generate Markov times $\{\tau_i\}$, a forecaster may, in addition, want to know the measure of the probability $P(\tau_i < \infty)_{i=1,2,\dots}$ before investing resources when applying this rule. In fact, if this probability is less than one, then the rule may never give a signal.

Definition 15. A Markov time τ is finite if $P(\tau < \infty) = 1$.

It turns out that only in very few cases the $\{\tau_i\}$ generated by technical analysis will be finite, and hence usable for practitioners. One exception is the method of moving averages.

Proposition 1. *If the observed process $\{X_t\}$ is stationary and m -dependent, all moving average methods characterized by (88) and (89) generate finite Markov times.*

Proof 1. *Let Z_t be given by (88). If X_t is stationary, then Z_t and Z_{t-1} are stationary. Let*

$$Y_t = Z_t Z_{t-1}, \quad t = 0, 1, \dots$$

Clearly $P(Y_t \geq 0) < 1$.

Equipped with proposition 6.38 of Breiman [5]:

$$P(Y_t \leq 0, \text{ at least once}) = 1,$$

then, the $\tau_1, \tau_2, \dots, \tau_n$, are almost surely and on the sample space $\{\omega : Y_0 \geq 0\}$ they form a stationary sequence under the probability $P(\cdot | Y_0 \geq 0)$, and

$$\mathbb{E}[\tau_1 | Y_0 \geq 0] = \frac{1}{P(Y_0 \geq 0)}.$$

We can consider the following:

$$P(Y_t \leq 0, \text{ at least once for } t \leq n) = 1 - P(Y_0 > 0, Y_1 > 0, \dots, Y_n > 0)$$

and

$$\begin{aligned} P(Y_0 > 0, Y_1 > 0, \dots, Y_n > 0) &= P(Y_0 > 0)P(Y_1 > 0) \dots P(Y_k > 0) \\ &= P(Y_0 > 0)^k \end{aligned}$$

letting $k \rightarrow \infty$,

$$P(Y_0 > 0)^k \rightarrow 0,$$

since $P(Y_0 > 0) < 1$, as shown above. Thus,

$$(104) \quad P(Y_n \leq \text{at least once}) = 1.$$

Hence, all conditions of the theorem supplied by proposition 6.38 of [5] are satisfied and Markov times $\tau_1, \tau_2, \dots, \tau_n$ are finite.

Chapter 3

Pairs Trading

Pairs trading is a strategy that tries to take advantage of market inefficiencies in order to obtain profit. The concept is disarmingly simple. First one must find two stocks whose price have moved together historically. Then ones takes long/short positions when they diverge abnormally, hoping that the prices will converge in the future. This sort of trade is regarded as SA, which only requires that the expected value of the trade be positive. A single trade has positive probability of loss but, as the number of trades increases, the Law of Large Numbers implies that the average returns approach their expected value. Since this expected value is positive, when the trading process is repeated many times, an arbitrage situation arises.

As stated in the literature review, there are some pairs trading models available. They range from a simplified approach such as the distance method, the co-integration approach or the stochastic spread and the stochastic residual spread in a discrete-time and continuous-time setting respectively.

The first feature that we would like to have in a pair is **co-integration**: two non-stationary time series have a unique linear combination that produce a stationary one (definition 10). As this kind of strategy relies on the fact that the spread will eventually converge to its long term value, we should require to do so with probability 1. Moreover, we would like to return there in finite time, that is equivalent to require that the models presents **ergodicity** as a main characteristic (definition 11). Namely, the long run average behavior mimics the instantaneous behavior. So an ergodic process will visit each state infinitely often and this implies the time between visits is certainly finite.

Another issue that a modeling framework of pairs trading should address is the construction of optimal trades in the mean-variance sense. In order to do so, one needs to calculate the expected return and variance of the trade in terms of the first-passage time. Finding the hitting time distributions is not trivial and in some cases no analytic formula is known.

In the following sections we review the methods available in both discrete-time and continuous time. Namely, the distance method, the co-integration method, the stochastic spread and the stochastic residual spread.

1. The distance method between vector prices

Let $P \in \mathbb{R}^{n \times m}$ denote the price matrix. This is, $p_{i,j} \in P$, the price of asset j at time i . Alternatively, the matrix P can be seen as a collection of column vectors $P = [p_1, \dots, p_m]$.¹

First, we need to bring all asset prices to a particular unit, and after that, to search stocks that move together. Let $\tilde{p} = \frac{p_{i,j} - \mu_j}{\sigma_j}$ be the transformation used to normalize $\log(P)$. Now we want to calculate the distance of a given asset with respect to the rest of the sample. Hence we need to calculate $m(m-1)/2$ distances.

There are several choices available to calculate the distance between vector prices. The most straight forward possibility is to chose the euclidean distance given by $d_{i,j} = (\|\tilde{\mathbf{p}}_i - \tilde{\mathbf{p}}_j\|^2)^{1/2}$. A less direct possibility, which involves the correlation coefficient $\rho_{i,j}$ (as seen in chapter 1), is to chose equation $d_{i,j} = \sqrt{2(1 - \rho_{i,j})}$ as a proper metric for measuring the distance between vectors.

The next step is to chose for each stock, a pair that has minimum squared distance between normalized prices. Once the plausible pairs have been identified, the trading rule triggers a signal every time that the absolute distance is higher than a specified parameter λ . There are several choices for λ . The most common decision is to chose the standard deviation as the trigger as in [17] or the α -quantile [39] of the empirical distribution of the spread.

In figure 20 we can appreciate the above examples of trading rules. Markov times are generated when the spread crosses the 2σ bound in the first case. In the second case we use the 1505 rule defined as follows:

¹ where $P(:, k) = \begin{bmatrix} p_{1k} \\ \vdots \\ p_{mk} \end{bmatrix}$

a trade is entered when the spread hits the 15th quantile and exits when it hits the 5th quantile. Opposite positions in the constituent pair are taken for the 65th and the 95th quantile respectively.

If we choose the euclidean distance as a metric for the spread, the smallest distance between the 435 pairs of the Dow Jones index is found to be the spread between AT&T and Verizon (two telecom services companies).

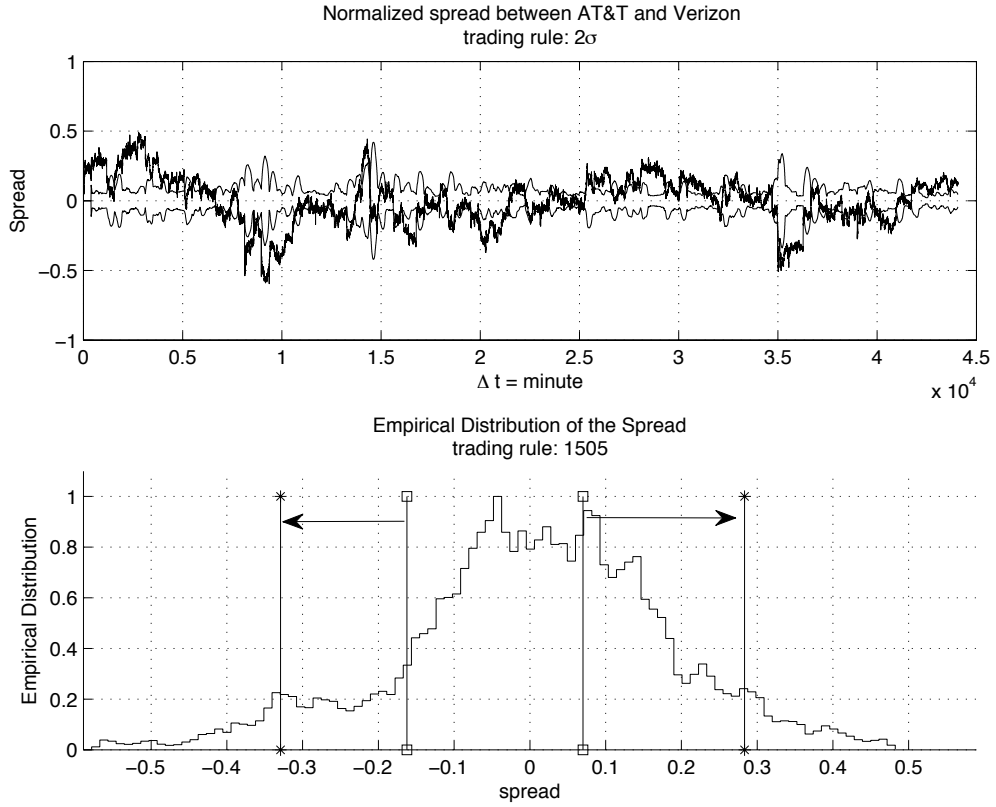


FIG. 20. Standard deviation rule (upper panel). α -quantile rule (lower panel). The squares represent entry points in long positions, conversely the stars represent the exit points.

The portfolio's P&L is calculated by means of:

$$(105) \quad \text{pnl}_i = (R)_{i,j} \cdot (I)_{i,j} + b \cdot T.$$

where,

$$(106) \quad R = \ln \begin{bmatrix} \frac{p_{1,1}}{p_{2,1}} & \cdots & \frac{p_{1,n}}{p_{2,n}} \\ \vdots & & \vdots \\ \frac{p_{m-1,1}}{p_{m,1}} & \cdots & \frac{p_{m-1,n}}{p_{m,n}} \end{bmatrix},$$

is the return matrix, $R \in \mathbb{R}^{(m-1) \times n}$.

$$(107) \quad I = \begin{bmatrix} I_{1,1} & \cdots & I_{1,n} \\ \vdots & & \vdots \\ I_{m-1,1} & \cdots & I_{m-1,n} \end{bmatrix},$$

is the indicator matrix, it takes -1 for a short position, 1 for a long position and 0 otherwise. $I \in \{-1, 0, 1\}^{(m-1) \times n}$.

$$(108) \quad T = \begin{bmatrix} T_{1,1} & \cdots & T_{1,n} \\ \vdots & & \vdots \\ T_{m-1,1} & \cdots & T_{m-1,n} \end{bmatrix},$$

is the transaction cost matrix, it takes the value 1 if a transaction was made and 0 otherwise. $T \in \{0, 1\}^{(m-1) \times n}$.

$$(109) \quad b = \ln \left(\frac{1 - c}{1 + c} \right).$$

is the transaction cost as a percentage of the transaction price c .

2. Co-integration

Although the relationship between co-integration and error correction models, was first suggested by Granger [19], these concepts were fully developed by Engle and Granger [13] and Johansen [26] who developed a theoretical framework for non stationary VAR(p) process. Even though two time series are non-stationary, it is possible that, in some instances, a specific linear combination of the two be stationary.

The explanation for co-integration dynamics is captured by the notion of error correction. What Engle and Granger proposed is that long-run components of variables obey equilibrium forces while short-run components have a flexible dynamic specification.

2.1. Vector autoregressive process

Consider the p -dimensional autoregressive process X_t defined by the equation:

$$(110) \quad X_t = A_1 x_{t-1} + A_2 x_{t-2} \cdots + A_p x_{t-p} + D s_t + \varepsilon_t.$$

where $X_t = (x_{1,t} \dots x_{n,t})$ is a multivariate stochastic time series, $A_i, i = 1, 2, \dots, p$ and D are deterministic $n \times n$ matrices. $\varepsilon_t = (\varepsilon_{1,t}, \dots, \varepsilon_{n,t})$ is a multivariate white noise with variance-covariance matrix $\Omega = \{\sigma_{i,j}\}$ and $s_t = (s_{1,t} \dots s_{n,t})$ is a vector of deterministic terms.

Using the lag-operator L ² notation, a VAR model can be written in the following form:

$$(111) \quad X_t = (A_1 L + A_2 L^2 + \cdots + A_n L^n) x_t + D s_t + \varepsilon_t.$$

VAR models can be written in terms of the differences ΔX_t in the following error correction form:

$$(112) \quad \Delta X_t = (\Phi_1 L + \Phi_2 L^2 + \cdots + \Phi_{n-1} L^{n-1}) \Delta x_t + \Pi L^{n-1} x_t + D s_t + \varepsilon_t,$$

where the first $n - 1$ terms are in first differences and the last term is in levels.

The stationarity and stability properties of a VAR model depend on the roots of the polynomial matrix:

$$(113) \quad I - A_1 z + A_2 z^2 + \cdots + A_n z^n = 0.$$

If all the roots of are strictly outside the unit circle, the VAR process is stationary. In this case, the VAR process can be inverted and rewritten as an infinite moving average of a white noise process.

Definition 16. *The components of a vector X_t are co-integrated of order d, b , denoted $X_t \sim CI(d, b)$ if:*

- i) all components of X_t are integrated of order d ³, $I(d)$.*
- ii) there exists a vector $\alpha (\neq 0)$ so that $Z_t = \alpha^T X_t \sim I(d-b)$, $b > 0$.*

Given n time series, there can be, from none to at most $n - 1$ co-integrating relationships. The vector α is not unique. There may be $r < n$ linearly independent vectors $(\alpha_1, \alpha_2, \dots, \alpha_r)$. If there were

²The lag-operator is useful for notational and algebraic convenience only. It is defined as follows: $LX_t = X_{t-1}$. Then the following use of the operator will be clear: $L^2 X_t = L(LX_t) = LX_{t-1} = X_{t-2}$. It is also known in some books and journals B (backward-shift operator).

³ A variable is integrated of order d if it can be transformed into a stationary series differentiating d times.

n independent vectors, X_t is itself stationary and the concept of a co-integration has no meaning. The vector $\alpha^T X_t$ is an r -dimensional vector of trend-stationary variables. It can be noticed that this definition is symmetric in the variables, i.e: there is no designated left-hand side variable. Also, the α vectors are identified up to a scale, since $\alpha_i x_t$ stationarity implies that $\alpha_i x_t$ is stationary.

2.2. Estimating co-integrated systems

2.2.1. *The Engle-Granger's two step method.* Here we will follow the two step estimator procedure proposed in [13]. In the first step the parameters of the co-integrating vector are estimated and in the second step they are used in the error correction form. Both steps require only a single equation least square estimation.

Once a test is performed to each time series in order to check that the order of integration of both are $I(1)$ ⁴, we proceed in the following way:

- (1) Run a static regression:

$$y_t = \beta^T x_t + \varepsilon_t,$$

where x_t is one- or higher-dimensional. Since the asymptotic distribution of β is not standard, Engle and Granger suggested to estimate $\hat{\beta}$ by an OLS and the test for unit roots in the following equation:

$$\hat{\varepsilon}_t = y_t - \beta^T x_t.$$

- (2) The second step is to estimate the long-run equilibrium relationship between the time series. The error correction models reads:

$$\Delta y_t = A_1(L)\Delta y_{t-1} + A_2(L)\Delta x_{t-1} + \gamma z_t + \varepsilon_t,$$

where $z_t = \hat{\varepsilon}_t$.

Estimation of these parameters can be done with OLS. They converge faster towards their parameter values in the presence of a co-integrating relationship. The ADF⁵ test can be used to determine the stationarity of the residuals series ε_t .

The Engle and Granger (E-G) approach is relatively straight forward and easily to implement in practice. However it present some drawbacks. They can be summarized as follows:

⁴If each series have different orders of integration, it can be said that they are not co-integrated.

⁵Augmented Dickey-Fuller statistic.

- (1) Order matters. E-G uses the residuals from either of the two equilibrium equations;

$$\begin{aligned} y_t &= \alpha_y + \beta x_t + \varepsilon_{1,t}. \\ x_t &= \alpha_x + \beta y_t + \varepsilon_{2,t}. \end{aligned}$$

This could lead to different results. Asymptotically a test for a unit root in $\varepsilon_{1,t}$ is equivalent to that for $\varepsilon_{2,t}$. This is not applicable to smaller sample sizes.

- (2) The very nature of a two-step estimator involves carrying forward the errors introduced in the first step to the second step. This can be seen in the ADF equation:

$$(114) \quad \Delta \hat{\varepsilon}_t = \alpha + \beta t + \delta \hat{\varepsilon}_{t-1} + \sum_{i=1}^p \delta_i \hat{\varepsilon}_{t-1}.$$

The coefficients are obtained regressing the residuals from another regression. The errors induced in the first step are carried to the second step.

2.2.2. The Johansen's method. The best way of testing for units roots is using the system maximum likelihood estimator of Johansen. This estimator also gives the asymptotically efficient estimates of the co-integrating vectors and of the adjustments parameters. The Johansen Method (JM) is the maximum likelihood estimator of the so-called reduced rank model. A detailed exposition of the method can be found in Johansen [26] and Hamilton [20]. However here we outline the general idea of the procedure.

- (1) Calculate auxiliary regressions.

The first step is to estimate a (p-1)th-order VAR; that is, regress the scalar $\Delta y_{i,t}$ on a constant and all the elements of the vectors $\Delta y_{t-1}, \Delta y_{t-2}, \dots, \Delta y_{t-p+1}$ by OLS,

$$(115) \quad \Delta y_t = \hat{\pi}_0 + \hat{\Pi}_1 \Delta y_{t-1} + \hat{\Pi}_2 \Delta y_{t-2} + \dots + \hat{\Pi}_{p-1} \Delta y_{t-p+1} + \hat{\mathbf{u}}_t.$$

After collecting the residuals $\hat{\mathbf{u}}_t$ of the OLS regression in vector form, we perform a second battery of regressions, by regressing the scalar y_{t-1} on a constant and $\Delta y_{t-1}, \Delta y_{t-2}, \dots, \Delta y_{t-p+1}$. We form $\hat{\mathbf{v}}_t$, a vector of residuals from this second battery of regressions.

$$(116) \quad y_t = \hat{\theta} + \hat{\Theta}_1 y_{t-1} + \hat{\Theta}_2 y_{t-2} + \dots + \hat{\Theta}_{p-1} y_{t-p+1} + \hat{\mathbf{v}}_t.$$

- (2) Calculate canonical correlations.

Next we calculate the sample variance-covariance matrices of the OLS residuals, $\hat{\mathbf{u}}_t$ and $\hat{\mathbf{v}}_t$:

$$(117) \quad \hat{\Sigma}_{vv} \equiv (1/T) \sum_{t=1}^T \hat{\mathbf{v}}_t \hat{\mathbf{v}}_t^T.$$

$$(118) \quad \hat{\Sigma}_{uu} \equiv (1/T) \sum_{t=1}^T \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t^T.$$

$$(119) \quad \hat{\Sigma}_{uv} \equiv (1/T) \sum_{t=1}^T \hat{\mathbf{u}}_t \hat{\mathbf{v}}_t^T.$$

$$(120) \quad \hat{\Sigma}_{vu} \equiv \hat{\Sigma}_{uv}^T.$$

From these, find the eigenvalues of the matrix:

$$(121) \quad \hat{\Sigma}_{vv}^{-1} \hat{\Sigma}_{uv} \hat{\Sigma}_{uu}^{-1} \hat{\Sigma}_{vu},$$

with the eigenvalues ordered $\lambda_1 > \lambda_2 > \dots > \lambda_n$. The maximum value attained by the log likelihood function subject to the constraint that there are r co-integrating relations is given by:

$$(122) \quad \begin{aligned} \mathcal{L}^* = & -(Tn/2) \log(2\pi) - (Tn/2) - (T/2) \log |\hat{\Sigma}_{uu}| \\ & - (T/2) \sum_{i=1}^h \log(1 - \lambda_i). \end{aligned}$$

(3) Calculate maximum likelihood estimates of parameters.

Denote $\hat{\mathbf{a}}_1, \dots, \hat{\mathbf{a}}_r$ the $(n \times 1)$ eigenvectors of (121) associated with the r largest eigenvalues. These provide a basis for the space of co-integrating relations; that is, the maximum likelihood estimate is that any co-integrating vector can be written in the form:

$$(123) \quad \mathbf{a} = b_1 \hat{\mathbf{a}}_1 + b_2 \hat{\mathbf{a}}_2 + \dots, b_r \hat{\mathbf{a}}_r.$$

Johansen suggested normalizing the $\hat{\mathbf{a}}_i$ so that $\hat{\mathbf{a}}_i^T \hat{\Sigma}_{vv} \hat{\mathbf{a}}_i$ are orthonormal.

Collect the first h normalized vectors in a $(n \times h)$ matrix $\hat{\mathbf{A}}$:

$$(124) \quad \hat{\mathbf{A}} \equiv [\hat{\mathbf{a}}_1 \quad \hat{\mathbf{a}}_2 \quad \dots \quad \hat{\mathbf{a}}_h].$$

Finally, the close form MLE is given by:

$$(125) \quad \mathcal{L}_0 = \hat{\Sigma}_{\mathbf{U}\mathbf{V}} \hat{\mathbf{A}} \hat{\mathbf{A}}^T.$$

$$(126) \quad \mathcal{L}_i = \hat{P}i_i - \mathcal{L}_0 \hat{\Theta}_i, \quad i = 1, 2, \dots, p-1.$$

2.3. Filtering OLS estimators

In this section we use a linear Gaussian model, the KF, to filter the time-varying estimation of the OLS in the E-G two-step estimation procedure.

The time varying beta model can be expressed by the following system of state-space equations:

$$(127) \quad y_t = \beta_t x_t + \varepsilon_t,$$

$$(128) \quad \beta_t = \beta_{t-1} + \eta_t,$$

where ε_t and η_t are independent uncorrelated error terms with mean $\langle \varepsilon_t \rangle = 0$, $\langle \eta_t \rangle = 0$ and variance H_t , Q_t respectively.

The equation (127) is known as a *Measurement equation* and (128) as the *State equation*, which in our case defines β as a simple random walk process.

The KF estimate for β is as follows:

$$(129) \quad \beta_{t|t-1} = \beta_t.$$

$$(130) \quad P_{t|t-1} = P_t + Q.$$

$$(131) \quad v_t = y_t - \beta_{t|t-1} x_t.$$

$$(132) \quad F_t = \beta_t P_{t|t-1} \beta_t^T + H.$$

$$(133) \quad x_t = x_{t|t-1} + \frac{P_{t|t-1} \beta_t^T v_t}{F_t}.$$

$$(134) \quad P_t = P_{t|t-1} - \frac{P_{t|t-1} \beta_t^T \beta_t P_{t|t-1}}{F_t}.$$

2.4. Model implementation

Let x_t and y_t be two non-stationary time series, if for certain value α , $\ln(y_t) = \beta \ln(x_t)$ is stationary, then the two series are said to be co-integrated. Co-integrated systems have a long-run equilibrium; that is, the long-run mean of the linear combination of the two series. If there is a deviation from the long-run mean, then one or both time series adjust themselves to restore the equilibrium.

The error correction representation is:

$$(135) \quad y_t - y_{t-1} = \alpha_y (y_{t-1} - \beta x_{t-1}) + \varepsilon_y,$$

$$(136) \quad x_t - x_{t-1} = \alpha_x (x_{t-1} - \beta x_{t-1}) + \varepsilon_x,$$

where, α is the error correction rate, indicative of the speed with which the time series corrects itself to maintain equilibrium and β is the coefficient of co-integration. The parameters $\Phi = \{\alpha_x, \alpha_y, \beta\}$, uniquely determine the model.

The decision process is as follows. Consider N days of history. Calculate and estimate the system with either the E-G or the Johansen's methodology. If we accept the null H_0 ⁶ the two series are not co-integrated, hence the process keep searching till a co-integrated sample is found. If we accept the null hypothesis, then the sample series are co-integrated. Thus, we proceed to check the residual of the regression. If the last component of the residual vector is found to be higher (lower) than its standard deviation we short (long) the regressand (endogenous variable) and long (short) the regressor (exogenous variable).

$$\text{Rule} \begin{cases} \text{if residual}(N) > \lambda \sigma^{\text{res}} \Rightarrow & -1 \cdot y_t + 1 \cdot x_t, \\ \text{if residual}(N) < \lambda \sigma^{\text{res}} \Rightarrow & 1 \cdot y_t - 1 \cdot x_t, \end{cases}$$

where λ is typically chosen to be 1 or 1.5 times σ .

3. Stochastic spread

SA exploits mathematical models to generate returns from systematic movements in securities prices. This encompasses a variety of strategies and investment programs including systematic trading signals in order to get a market-neutral trading book. Pairs trading is considered an “ancestor” of SA. It is based on identifying pairs of securities that typically trade in a predictable relation to one another. In this section we will discuss an application of the concepts detailed in section 5.5 of chapter 1 for modeling the spread in continuous time. The aim is to specify an analytical tractable model, and to be able to calibrate the model with market data.

3.1. Univariate spread dynamics

Here we recall the basic spread model outlined in chapter 1 and formulate a pairs trading stochastic spread strategy.

Assuming that,

$$(137) \quad x_t = \log(P_t^A) - \log(P_t^B),$$

is governed by an Ornstein-Uhlenbeck (O-U) process:

$$(138) \quad dx_t = \theta(\mu - x_t)dt + \sigma dW_t,$$

where,

- θ : speed or reversion.

⁶The sample time series is a random walk.

- μ : long term mean.
- σ : volatility.
- dW : Wiener process.

Its close-form solution reads

$$(139) \quad x_t = \mu(1 - e^{-\theta t}) + e^{-\theta t}x_0 + \int_0^t \sigma e^{\theta(s-t)} dW_s.$$

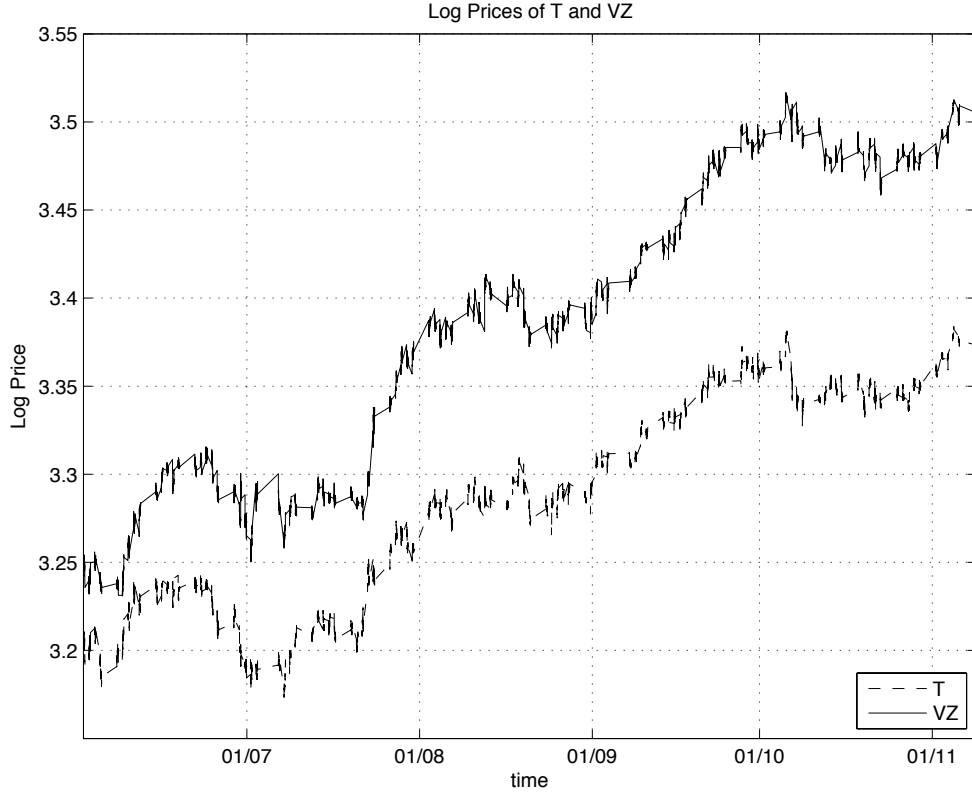


FIG. 21. Log prices of AT&T and Verizon.

From (139), and assuming that x_0 is a constant, we can extract the first moment directly: $E[x_t] = \mu(1 - e^{-\theta t}) + (e^{-\theta t})x_0$. The integral in the right hand side is normally distributed with mean zero and variance $\mathbb{E}\left[\left(\int_0^t e^{\theta(s-t)} \sigma dW_s\right)^2\right]$. Using the Itô isometry⁷ we get the variance of the process:

$$\bullet \text{Var}[x_t] = \frac{\sigma^2}{2\theta}(1 - e^{-2\theta t}),$$

where the long term ($t \rightarrow \infty$) mean and variance are:

$$\bullet E[x_t] = \mu.$$

⁷see subsection 5.3 of chapter 1

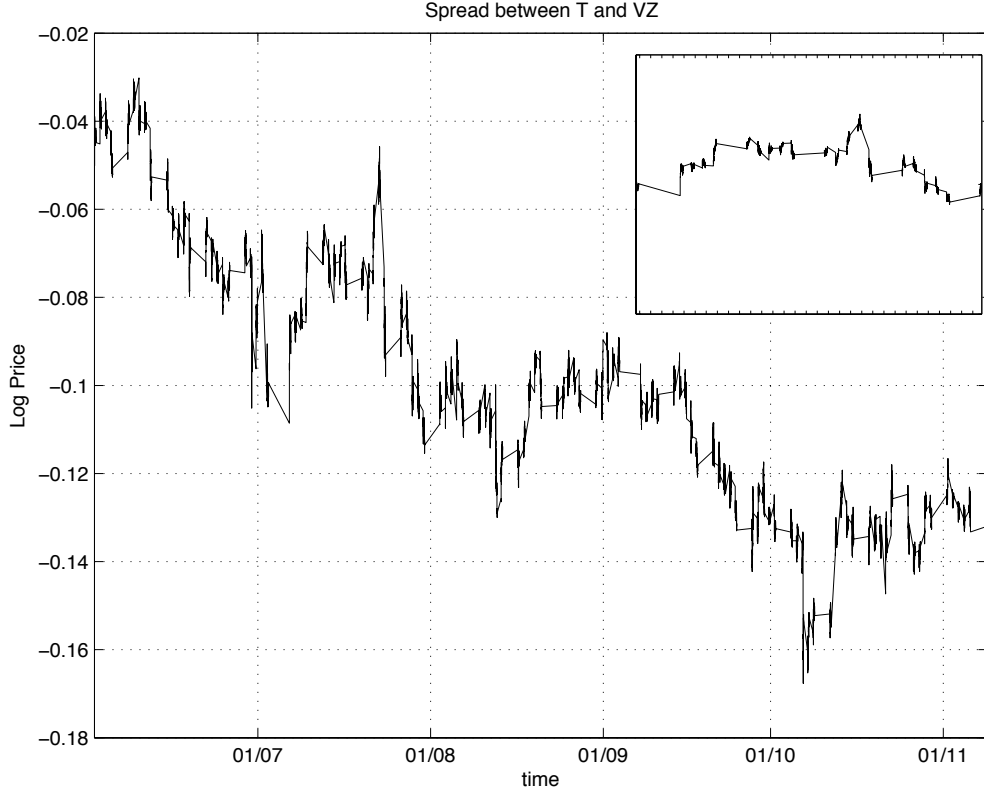


FIG. 22. Spread of AT&T and Verizon. Inset: A spread roundtrip from 2-jul-2010 to 2-ago-2010.

- $Var[x_t] = \frac{\sigma^2}{2\theta}$.

The mean reverting behavior of the SDE can be studied through a feature of its discrete-time counterpart, the autoregressive process. Since for the AR(1) process if $0 < \beta < 1$ it is a necessary and sufficient condition for stationarity, and hence mean reversion. The case $\beta = 1$ corresponds to a random walk.

In order to so, we first discretize (139):

$$(140) \quad x_t = \alpha + \beta x_{t-1} + \varepsilon_t,$$

where,

- $\alpha = \mu(1 - e^{-\theta\Delta t})$.
- $\beta = e^{-\theta\Delta t}$.
- $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$.
- $\sigma_\varepsilon^2 = \frac{\sigma^2}{2\theta}(1 - e^{-2\theta\Delta t})$.

To fit an specific mean reverting process to data, it is important to check the validity of the mean reverting assumption. A simple way to

do this is to test stationarity. For a discussion of tests available see [6]. Here we will use the ADF statistic.

The presence of a unit root indicates that the time series is not stationary, but differencing will reduce it to stationarity. See [10] for a detailed explanation on unit root testing. Rejecting the null hypothesis of the ADF test $H_0 : \beta = 1$ will confirm stationarity, mean reversion and hence, allow us to use it in our trading strategy. Conversely, if we accept the null, no trading will take place. There are several statistical and computing packages with ready to use tests.⁸

3.1.1. Parameter estimation. To calibrate the discrete form of the process, the coefficients α , β and $\delta = \text{std}(\text{residuals})$, are estimated using an OLS regression of the time series x_t on its lagged form x_{t-1} .

Rewriting the systems of equations in (140) one gets the following parameters:

$$(141) \quad \begin{cases} \theta &= -\log(\beta)/\Delta t. \\ \mu &= \alpha/(1 - \beta). \\ \sigma &= \delta/\sqrt{(\beta^2 - 1)\Delta/2 \log(\beta)}. \end{cases}$$

Before doing so, an ADF test is performed on the residuals, as mentioned in the previous section, if we reject the null (the process is not a random walk), we estimate the coefficients of an AR(1), on the contrary, we keep looking in our time series till a stationary series appears. Thus the trading rule, once the mean reversion feature has been accepted, would be:

$$\text{Rule} \begin{cases} \text{if } x_t < \mu \Rightarrow -1 \cdot x_t + 1 \cdot y_t. \\ \text{if } x_t > \mu \Rightarrow -1 \cdot y_t + 1 \cdot x_t. \end{cases}$$

3.1.2. State space formulation. The equation (140) can be recast in a state space model in order apply a filter to estimate the parameters:

$$(142) \quad x_t = \alpha + \beta x_{t-1} + \gamma \varepsilon_t,$$

$$(143) \quad y_{t-1} = x_{t-1} + \delta \omega_{t-1},$$

with (142) and (143) as state equation and observation equation respectively. We need to estimate $\vartheta \equiv \{\alpha, \beta, \gamma, \delta\}$ or rather $\vartheta \equiv \{\alpha, \beta, \gamma^2, \delta^2\}$ from the observed data using the Kalman Filter.

We set $\mathcal{F}_t = \sigma\{y_0, y_1, \dots, y_t\}$ which represents the information from observing y_0, y_1, \dots, y_t . The objective is to compute the conditional expectation filter $\hat{x}_t = \mathbb{E}[x_t | \mathcal{F}_t]$, which are the best estimates of the

⁸SAS, R, Matlab, etc.

hidden state process. Given ϑ , and using the Kalman Filter, we can compute $\hat{x}_t \equiv \hat{x}_{t|t} = \mathbb{E}[x_t|\mathcal{F}_t]$ and $R_t = \Sigma_{t|t} \equiv \mathbb{E}[(x_t - \hat{x}_t)^2|\mathcal{F}_t]$.

Then (\hat{x}_t, R_t) are determined recursively as follows:

$$\begin{array}{rcl}
 \hat{x}_{t+1|t} & = & \alpha + \beta \hat{x}_{t|t} \\
 \Sigma_{t+1|t} & = & \beta^2 \Sigma_{t|t} + \delta^2 \\
 \mathcal{K}_{t+1} & = & \Sigma_{t+1|t} / (\Sigma_{t+1|t} + \delta^2) \\
 \hat{x}_{t+1|t+1} & = & \hat{x}_{t+1|t} + \mathcal{K}_{t+1} [y_{t+1} - \hat{x}_{t+1|t}] \\
 \Sigma_{t+1|t+1} & = & \delta^2 \mathcal{K}_{t+1} = \Sigma_{t+1|t} - \mathcal{K}_{t+1} \Sigma_{t+1|t}
 \end{array}$$

TABLE 5. Recursive equations for the KF.

For initialization one could take $\hat{x}_0 = y_0$ and $\Sigma_{0|0} = \delta^2$.

Let

$$(144) \quad \mathcal{L}_t(\vartheta) = \mathbb{E}_0 \left[\frac{dP_\vartheta}{dP_0} | \mathbb{F}_t \right],$$

be the likelihood function for $\vartheta \in \Theta$. The maximum likelihood estimates solves:

$$(145) \quad \hat{\vartheta} = \arg \max_{\vartheta \in \Theta} \mathcal{L}_t(\vartheta).$$

The Expectation-Maximization (EM) Algorithm is an iterative method to compute $\hat{\vartheta}$. Provided that $\hat{\vartheta}_0$ is given as an initial estimate, the EM-algorithm gives $\hat{\vartheta}_j, j = 1, 2, \dots$, as a sequence of estimates.

The first step (E-step) is to compute:

$$(146) \quad Q(\vartheta, \tilde{\vartheta}) = \mathbb{E}_{\tilde{\vartheta}} \left[\log \frac{dP_{\tilde{\vartheta}}}{dP_{\vartheta}} | \mathcal{F}_t \right]$$

with $\tilde{\vartheta} = \vartheta_j$.

The second step is as follows:

$$(147) \quad \vartheta_{j+1} \in \arg \max_{\vartheta \in \Theta} Q(\vartheta, \hat{\vartheta}_j).$$

The procedure to implement the EM-Algorithm used here is due to Shumway and Stoffer [46]. For a detailed exposition see [12].

3.2. Multivariate spread dynamics

Equation (138) can be written in matrix form to represent a multivariate O-U process:

$$(148) \quad d\mathbf{X}_t = \mathbf{\Theta}(\boldsymbol{\mu} - \mathbf{X}_t)dt + \boldsymbol{\sigma}d\mathbf{W}_t,$$

where $\mathbf{\Theta}$ is the transition matrix, namely a full generic square matrix that defines the deterministic portion of the evolution of the process. The unconditional expectation is given by a generic vector $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ defines the scatter generator. A vector of independent Wiener processes is given by \mathbf{W}_t .

Introducing the integrator:

$$(149) \quad \mathbf{Y}_t \equiv e^{\mathbf{\Theta}t}(\boldsymbol{\mu} - \mathbf{X}_t)$$

and using Itô's lemma we obtain:

$$(150) \quad d\mathbf{Y}_t = e^{\mathbf{\Theta}t}\boldsymbol{\sigma}d\mathbf{W}_t.$$

Finally, integrating both sides and substituting the definition (149) we obtain

$$(151) \quad \mathbf{X}_t = (\mathbf{I} - e^{\mathbf{\Theta}t})\boldsymbol{\mu} + e^{\mathbf{\Theta}t}\mathbf{X}_0 + \int_0^t \boldsymbol{\sigma}e^{\mathbf{\Theta}(s-t)}d\mathbf{W}_s.$$

Following [36] we can investigate a geometrical interpretation of (151). Consider the eigenvalues of the transition matrix $\mathbf{\Theta}$: since the matrix has real entries, its eigenvalues are either real or complex conjugate. We denote respectively by $(\lambda_1, \dots, \lambda_K)$ and $(\gamma_1 \pm i\omega_1), \dots, (\gamma_J \pm i\omega_J)$, where $K+2J=N$. Now consider the matrix \mathbf{B} whose columns are the respective, possibly complex, eigenvectors and define the real matrix $\mathbf{A} \equiv \text{Re}(\mathbf{B}) - \text{Im}(\mathbf{B})$. Then the transition matrix can be decomposed in terms of eigenvalues and eigenvectors as follows:

$$(152) \quad \mathbf{\Theta} \equiv \mathbf{A}\mathbf{\Gamma}\mathbf{A}^{-1},$$

where $\mathbf{\Gamma}$ is a block-diagonal matrix,

$$(153) \quad \mathbf{\Gamma} \equiv \text{diag}(\lambda_1, \dots, \lambda_K, \mathbf{\Gamma}_1, \dots, \mathbf{\Gamma}_J),$$

and the generic j-th matrix $\mathbf{\Gamma}_j$ is defined as:

$$(154) \quad \mathbf{\Gamma}_J \equiv \begin{pmatrix} \gamma_j & \omega_j \\ -\omega_j & \gamma_j \end{pmatrix}.$$

Introducing a new set of coordinates,

$$(155) \quad \mathbf{z} \equiv \mathbf{A}^{-1}(\boldsymbol{\mu} - \mathbf{X})$$

Then the original (148) reads:

$$(156) \quad d\mathbf{Z}_t = \boldsymbol{\Gamma}\mathbf{Z}_t dt + \mathbf{V}d\mathbf{W}_t,$$

where $\mathbf{V} \equiv \mathbf{A}^{-1}\boldsymbol{\sigma}$.

Since this is another O-U process, its solution is normal $\mathbf{Z}_t \sim \mathcal{N}(\mathbf{z}_t, \boldsymbol{\Phi}_t)$ for a suitable deterministic drift \mathbf{z}_t and covariance $\boldsymbol{\Phi}_t$. The deterministic drift is the solution of the ordinary differential equation:

$$(157) \quad d\mathbf{z}_t = \boldsymbol{\Gamma}\mathbf{z}_t dt.$$

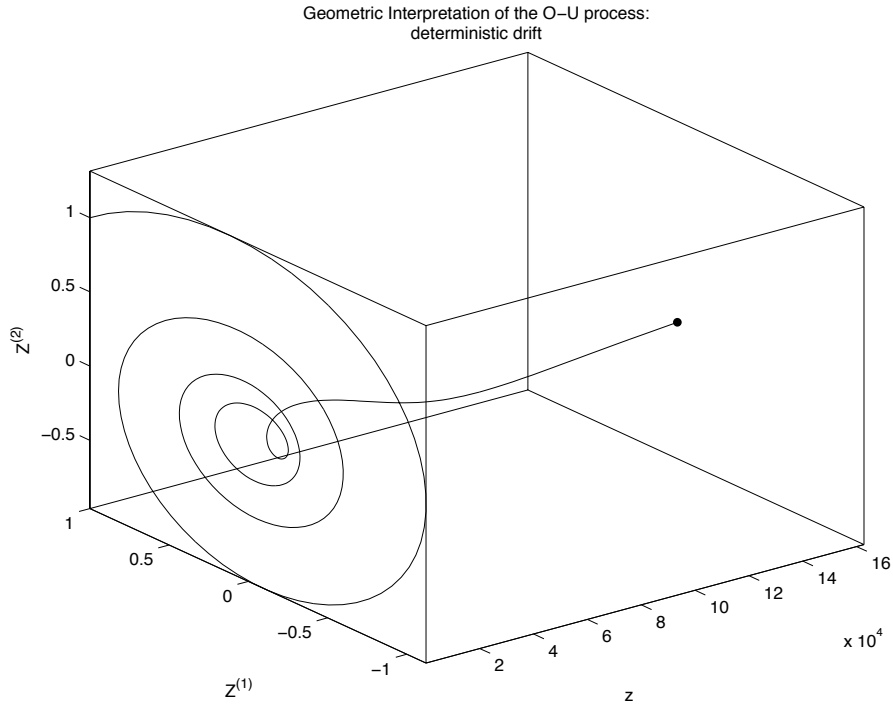


FIG. 23. Geometrical interpretation of the deterministic drift of the O-U process.

In figure 23 we can appreciate the exponential convergence of the three variables O-U process with unconditional mean $\boldsymbol{\mu} = [0 \ 0 \ 0]$. Conversely, for negative values of λ_k , the trajectory would be an exponential explosion.

4. Stochastic residual spread

So far, the existing methods reviewed address the pairs trading strategy issues based purely on a statistical basis. Do et al. in [11] have proposed a relative pricing model adapting the Arbitrage Pricing Theory (APT) model of a single asset.

The method of stochastic residual spread (SRS) proposed in [11] starts with the assumption that there exists an equilibrium in the relative valuation of the two stocks measured by some spread. The APT asserts that the return of a risky asset should be a linear combination of the risk premiums (above the risk free rate) and the exposures of the asset. However, it does not prescribe which factors one must use.

A “relative” APT on a pair of assets could be written as:

$$(158) \quad R_t^A - R_t^B = \Gamma R_t^m + \varepsilon_t,$$

where, $\Gamma = (\beta_1^A - \beta_1^B)$.

β is the risk factor.

R_t^m is the benchmark return. (i.e. S&P 500, DJIA).

$\varepsilon_t \sim \mathcal{N}(0, 1)$.

They specified a residual spread function as:

$$(159) \quad G_t = R_t^A - R_t^B - \Gamma R_t^m,$$

and reproduce the model in a state space setting proposing that the spread is governed by an O-U process:

$$(160) \quad dx_t = \theta(\mu - x_t)dt + \sigma dW_t,$$

as transition equation, and:

$$(161) \quad G_t = x_t + \omega_t,$$

as measurement equation.

Equation can be discretized in the same fashion as in section 3 of this chapter leading to the following discretized version of the above state space model, with the transition equation:

$$(162) \quad x_t = \mu(1 - e^{-\theta\Delta t}) + e^{-\theta\Delta t}x_{t-1} + \varepsilon_t,$$

where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ with $\sigma_\varepsilon^2 = \frac{\sigma^2}{2\theta}(1 - e^{-2\theta\Delta t})$.

And *Measurement equation* as:

$$(163) \quad y_t = x_t + \omega_t.$$

This state space model remains problematic with the observation G_t being still unobserved as Γ is unknown. One possibility is to perform a two-step estimation: first estimate Γ performing an OLS with $R_t^A - R_t^B$ as dependent variable and the excess return factors as regressors. The residual spread is constructed with the residuals from that regression. Second, this time series becomes the observation equation in the above state space model. In order to avoid doubling up the errors of the above methodology, the observation equation is redefined:

$$(164) \quad y_t = x_t + \Gamma r_t + \omega_t.$$

Equations (162) and (164) constitute a model of stochastic residual spread for a pairs trading implementation. This model is linear and Gaussian, which can be estimated by Maximum Likelihood Estimation (MLE) or some form of filtering as seen in section 7 of chapter 1.

In this strategy, the trading rule is based on a mispricing at the return level. The proposed strategy opens positions when the accumulated residual spread in the returns is sufficient large, and unwind when the accumulated spread is equal the long run level of the spread. To illustrate this point, consider stock A and B. Assuming that the last observed period sees A return 5% and B 3%, or a residual spread of 2%. For a correction to happen in the next period, the residual spread needs to be around -2%, regardless of the individual direction of the stocks, hence a zero accumulated residual spread.

Therefore, a trading rule for this strategy is to take a long-short position whenever the accumulated spread $\delta_k = \sum_{t=k-l}^k x_t$, with l less or equal to the current time k , exceeds μ . The trader will have to fix a base from which to determine the point l where $\delta_l = 0$.

$$\text{Rule} \begin{cases} \text{if } \sum_{t=k-l}^k x_t < \mu \Rightarrow -1 \cdot x_t + 1 \cdot y_t. \\ \text{if } \sum_{t=k-l}^k x_t > \mu \Rightarrow -1 \cdot y_t + 1 \cdot x_t. \end{cases}$$

Chapter 4

Numerical Results

The data series used in this study comes from the Dow Jones Industrial Average (DJIA), the same used in the empirical analysis conducted in section 6 of chapter 1. The data set spans from 01-Jun-2010 to 08-Nov-2010. The idea is to investigate whether the strategies outlined in chapters 2 and 3 generates positives P&L's. The testing is done splitting the data set in two: the first part is called the *in-sample* period where the models are trained and the second part is called the *out-of-the sample* period, where the model performance is scrutinized. The objective of the in-sample testing is to calibrate the optimal set of parameters. Optimality in a sense that a given set maximize a given measure of performance (i.e: Sharpe ratio, MM, etc.). The in-sample sample is characterized by two parameters: the breadth and the length. The first parameter deals with the size of the universe of the in-sample (i.e: U.S listed stocks, emerging market debt, etc.). The second aspect is simply the size of the window to train the model. The model simply has to do a reasonable job explaining the in-sample period in order to be further considered.

1. Testing a technical trading model

We start with a model from technical trading, more specifically a combo model outlined in section 6 of chapter 1. This model average two Markov times generated by an oscillator and a moving average. More specifically, the signals generated by the relative strength index and the exponential moving average.

From the practitioner perspective, there are a few questions that must be answered before committing capital and resources to a certain quantitative strategy. Namely, from the investment universe one has to pick a certain equity, and the trader must decide the size of the rolling windows and the sample frequency. Finally, the investor must choose its leverage or debt to capital ratio.

From a theoretical perspective we wish to explore whether the model is well behaved, that is, if it is very sensitive to the initial parameters and if the returns and results of a given sample can be extrapolated them.

Figure 24 it is useful from two different perspectives. For the practitioner it is extremely helpful to calibrate the model at hand, and execute the trades according to the parameters which gives the higher SR. From a modeling point of view we appreciate the instability of the model. It seems that a small shift from the “optimal” parameters affects considerably the output. For this reason, one should be extremely careful when using this models and “real-time” calibration is recommended.

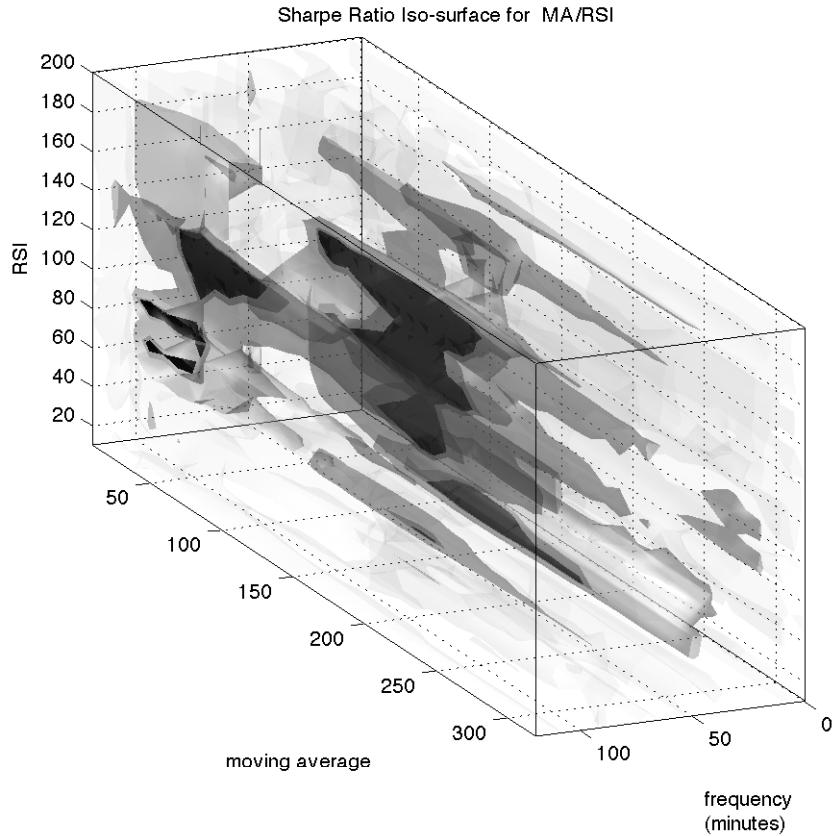


FIG. 24. Sharpe’s ratio iso-surface for a MA/RSI implementation for XOM. The darker areas represents higher SR.

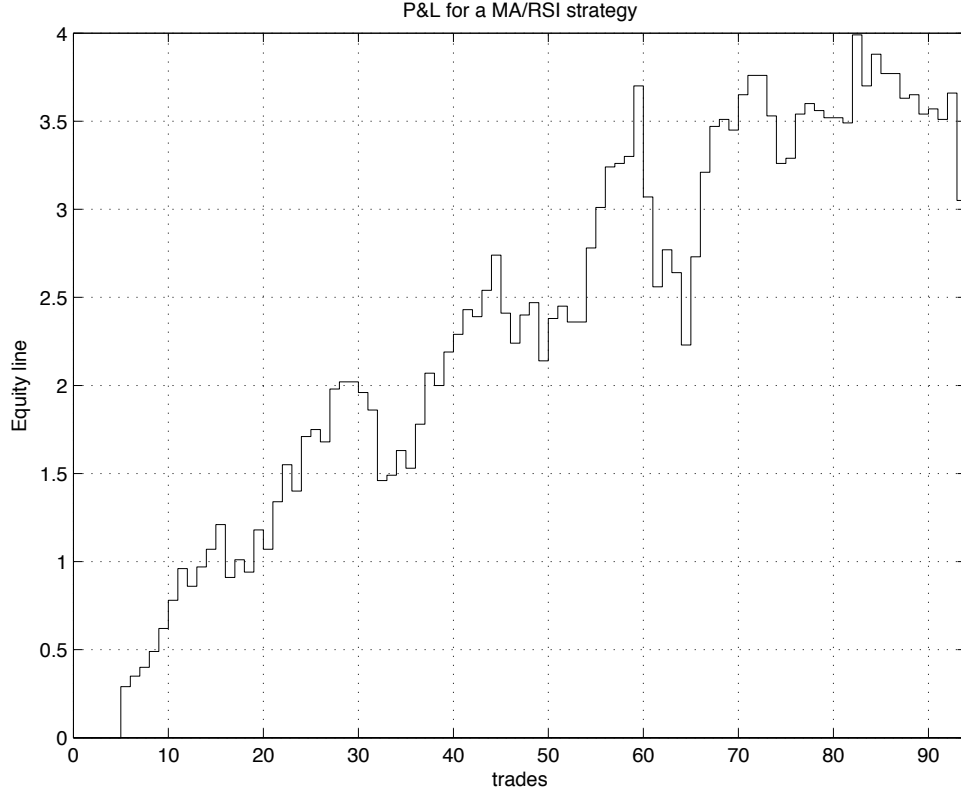


FIG. 25. P&L (out-of-the-sample) generated by a mixed model calibrated with $\Delta t = 70$, MA=160 and RSI=140.

Parameter	Optimal value
Frequency	70
Mov. Avg	160
RSI	140

TABLE 6. Optimal parameters calibrated in the in-sample data set (75% of the data points).

2. Testing pairs trading models

2.1. The distance strategy

The next implementation involves the most basic pairs trading strategy: the distance method with the standard deviation rule to generate the signals. The input of the model is a price matrix of the DJIA. The dimension of the matrix is $44,070 \times 30$. That is 113 trading days sampled minutely.

Strategy Summary	
Sharpe Ratio	4.2
Gross Profit	2.67
Total Trades	188
Largest Win	0.5
Largest Loss	-0.63
Max Drawdown	0.46
Max DD Duration	12
Avg. Profit per Trade	0.03
Std. per Trade	0.22

TABLE 7. Performance (out-of-the-sample) of the MA/RSI method calibrated with $\Delta t = 70$, MA=160 and RSI=140 for XOM.

As we mentioned before we are going to split the dataset in two. With the first section of the dataset, the in-sample (85 trading days¹, we are going to find the closest pair according to the squared distance criteria and the optimal window for calculating the spread and the optimal threshold, that is, how k-times the standard deviation is considered to trigger a signal. Although the constituent stocks of the index pertain to the most liquid securities, they are not perfectly synchronized. In order to homogenize the prices we take the most recent record to fill the missing data.

AT&T (T) and Verizon (VZ) were found to be the minimum distance pair to backtest the standard deviation rule.

The benchmark chosen for this strategy is an equally weighted portfolio. According to section 6.2 of chapter 1 this is the eigenvector of the largest eigenvalue of the correlation matrix of the DJIA. The minutely mean and standard deviation for 113 trading days (entire data set) for the benchmark is 0.03×10^{-4} and 5.16×10^{-4} respectively. Hence, 27.7% is the annualized return and 16.11% is the annualized standard deviation.² This is equal to a SR of 1.72 and a MM of 27.7%. If we only consider the out-of-the-sample period the same figures are: minutely mean and standard deviation, 0.04×10^{-4} and 4.25×10^{-4} respectively. Their annualized values are: 41% and 16%. Finally, the Sharpe ratio is 3.12 and, obviously the MM is the same as the annualized return.

¹A rule of thumb for backtesting pairs trading strategy: 75% percent of the data points as the in-sample size. That is 33,150 and 10,920 data points for the in-sample and out-of-the-sample periods respectively.

²To annualize returns, we consider 97,500 minutes in 250 days consisting of 6.5 hours of trading. To annualize the standard deviation one has to multiply the minutely standard deviation by $\sqrt{97,500} = 312.25$.

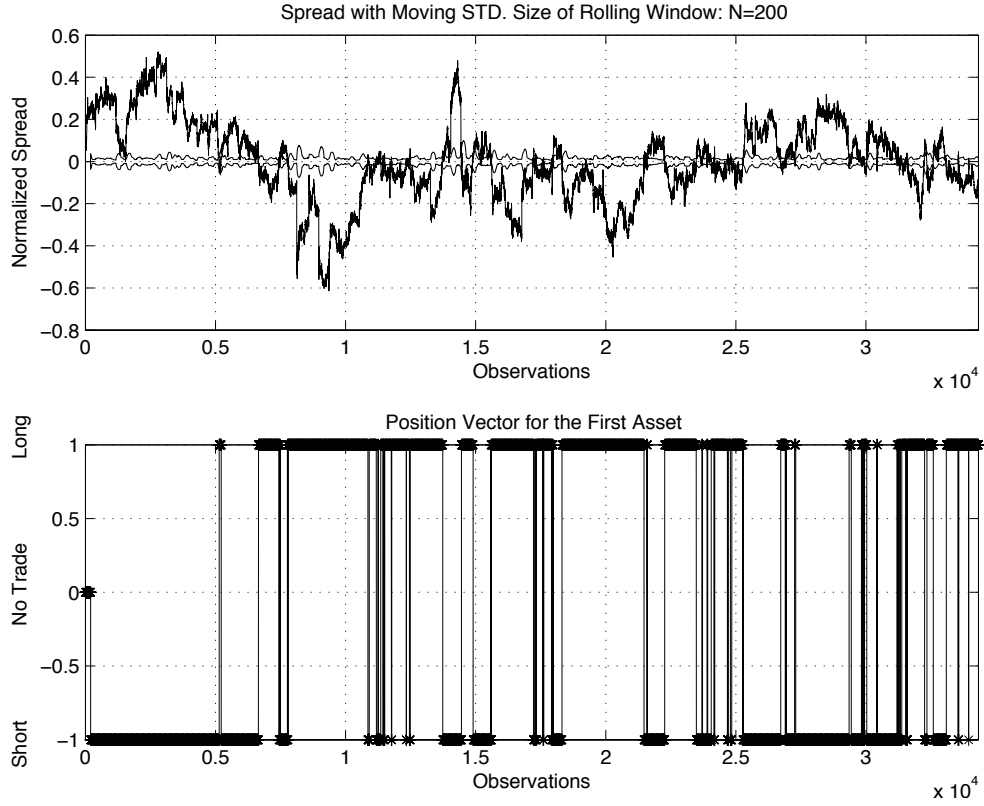


FIG. 26. T and VZ spread (in-sample) with rolling standard deviation (upper panel). Position vector for the first asset (lower panel).

Strategy Summary	
Sharpe Ratio	8.93
MM	118.62%
Gross Profit	5.03
Total Trades	60×2
Largest Win	0.47
Largest Loss	-0.13
Max Drawdown	1.56
Max DD Duration	2698(minutes)
Avg. Profit per Trade	0.09
Std. per Trade	0.09

TABLE 8. Performance (out-of-the-sample) of the distance method for T and VZ calibrated with a rolling window of 140 minutes.

This strategy outlined in table 8 clearly outperforms the unmanaged portfolio chosen as benchmark.

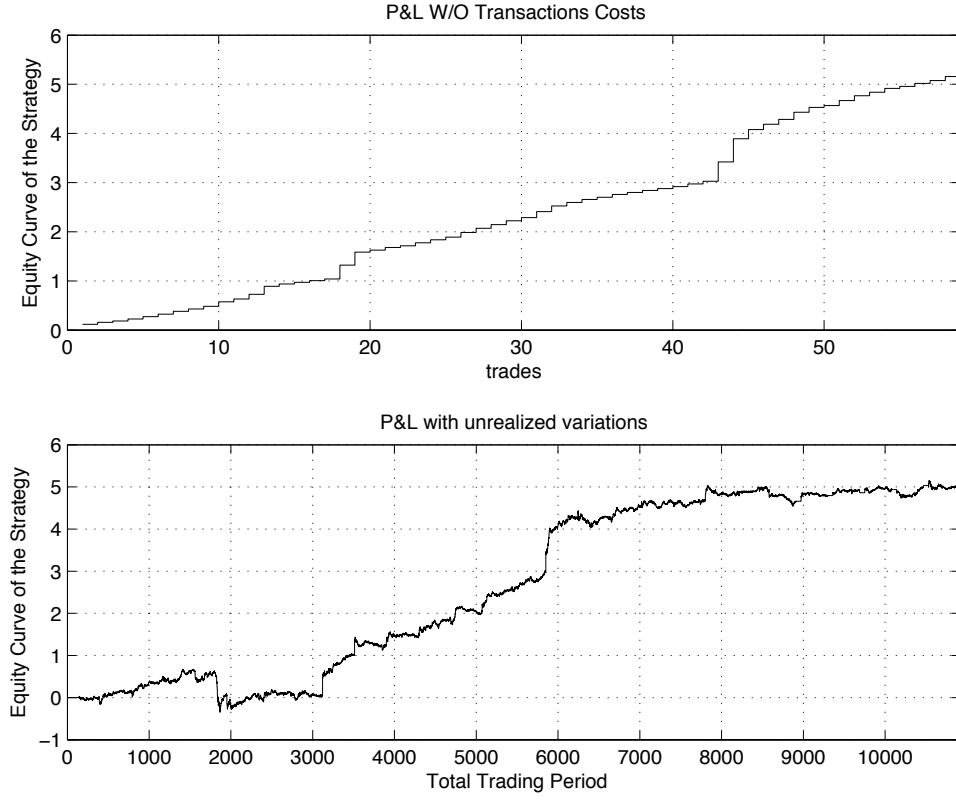


FIG. 27. P&L generated (out-of-the-sample) by the distance method with the standard deviation rule.

2.2. The co-integration strategy

The objective now is to backtest the co-integration strategy with market data. First, we check the ADF statistic to see whether there are co-integrated pairs in the DJIA or not. Figure 28 tell us the likelihood of co-integration of each of the 435 pairs. The darker the pair the most likely to be co-integrated.

HPQ and DIS³ are found to be, according to the ADF statistic, a good candidate to test the co-integration strategy. The coordinates of the chosen pair on figure 28 are (14,30).

After choosing the pair to back-test the co-integration strategy we perform an in-sample calibration of the size of the rolling window to trade according to the rules of this method.

In figure 29 we appreciate that this method does not behave well in the in-sample test. That is, a small perturbation in the parameters leads to a big change in the output measure, in this case the Sharpe ratio.

³See table A.

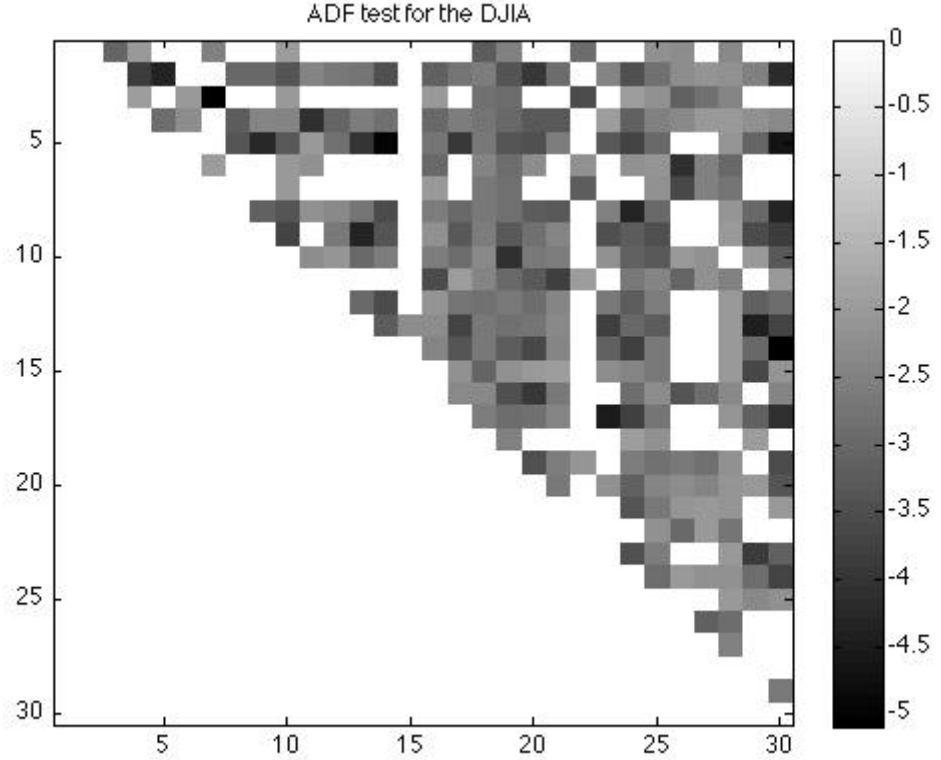


FIG. 28. Heat Map: ADF test for the DJIA price matrix. The darker squares represents likelihood of co-integrability.

In our case the best in-sample rolling window to trade is found to be 250 minutes and recalibrate every 25 minutes.

Table 9 describes some statistics of the output of the co-integration strategy for the out-of-the-sample period. Although the test outperformed the benchmark, it is by no means promising. The maximum drawdown duration is about 19 days and the standard deviation of each trades is one order of magnitude bigger than the average profit.

2.3. The stochastic spread strategy

Now we turn to test the stochastic spread strategy. We choose the pair with the shortest squared distance as in the distance method outlined before in this chapter. That is: AT&T and Verizon. As usual we calibrate the in-sample rolling window (75% of the data set) to maximize the Sharpe ratio. In this case this method seems to be well behave as shown in figure 31. This strategy performs better at smaller sample frequency, 25 minutes, and it needs also a recalibration every 5 minutes to achieve the maximum performance in the in-sample data set.

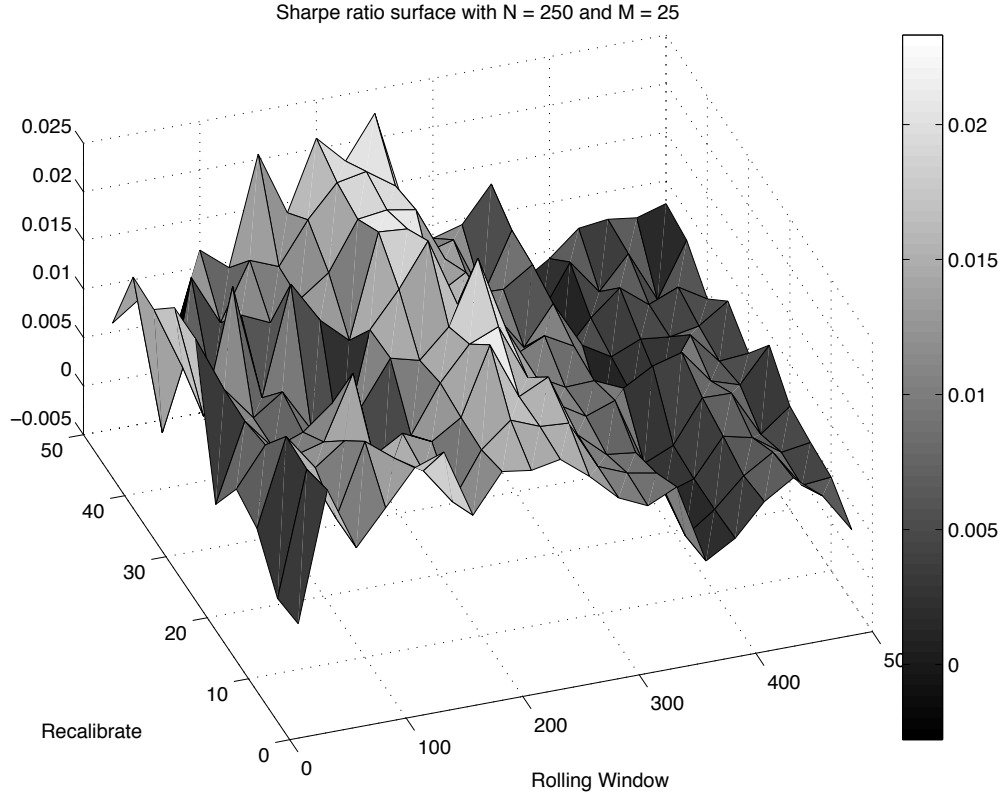


FIG. 29. Sharpe ratio surface for the co-integration method (in-sample).

Strategy Summary	
Sharpe Ratio	3.66
MM	48.67%
Gross Profit	3.88
Total Trades	406×2
Largest Win	1.14
Largest Loss	-0.52
Max Drawdown	0.98
Max DD Duration	7495 (minutes)
Avg. Profit per Trade	0.01
Std. per Trade	0.11

TABLE 9. Performance (out-of-the-sample) of the co-integration method calibrated with a rolling window of 250 minutes for HPQ and DIS.

The out-of-the-sample statistics (see table 10) are not compelling. The gross profit is half as the co-integration method and a third of the

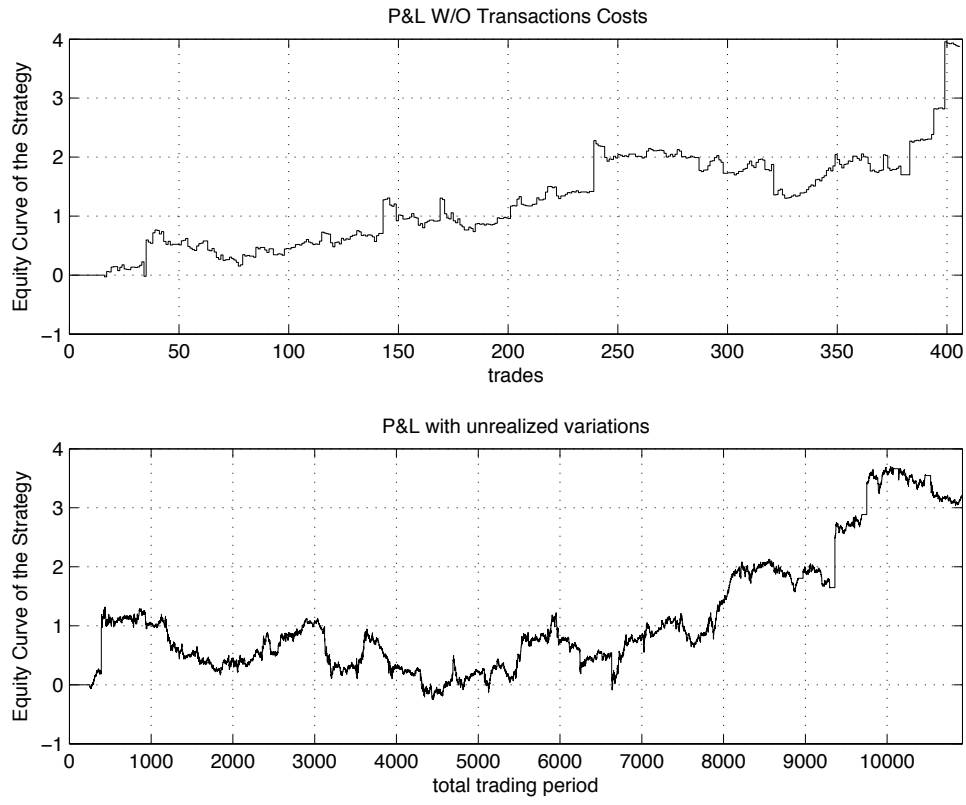


FIG. 30. P&L generated (out-of-the-sample) by the co-integration method for HPQ and DIS.

distance method. Although the Sharpe ratio and MM ratio are sufficient, when including transaction cost in the backtesting the gross profit would be wiped out.

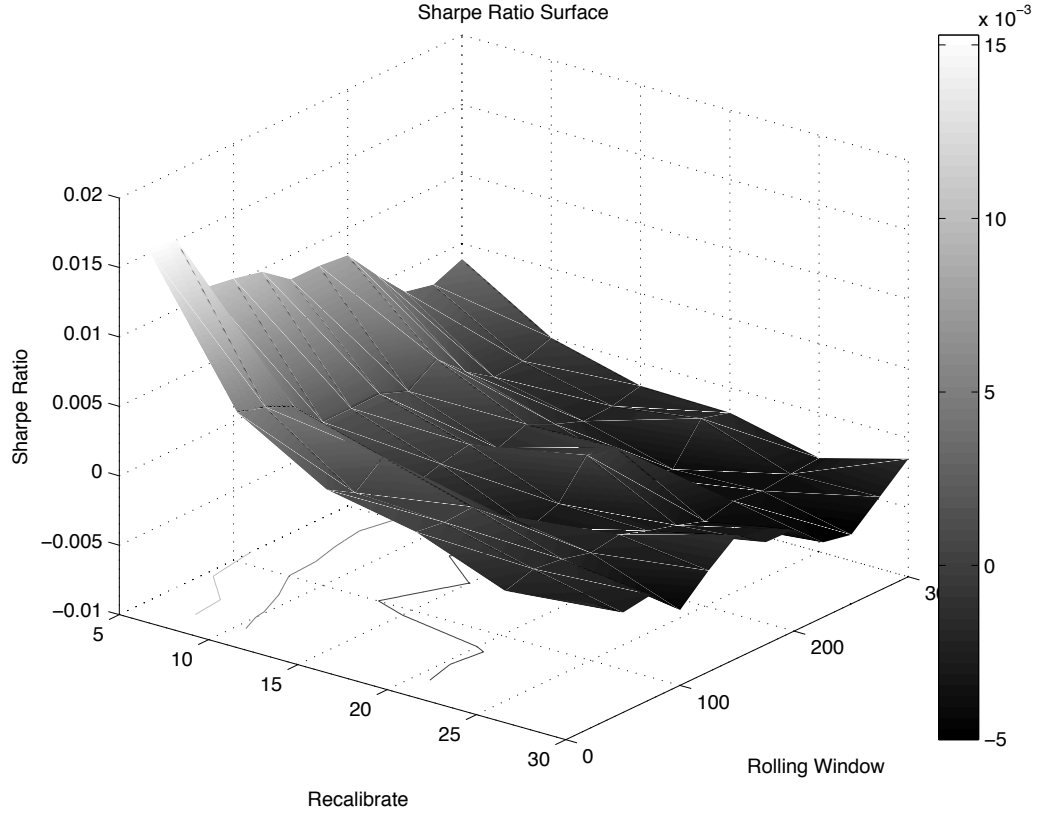


FIG. 31. Sharpe Ratio Surface (in-sample) for the stochastic spread method.

Strategy Summary	
Sharpe Ratio	4.99
MM	66.21%
Gross Profit	1.56
Total Trades	2226×2
Largest Win	0.54
Largest Loss	-0.61
Max Drawdown	1.30
Max DD Duration	2738 (minutes)
Avg. Profit per Trade	0.07×10^{-2}
Std. per Trade	0.02

TABLE 10. Performance (out-of-the-sample) of the stochastic spread method calibrated with a rolling window of 25 minutes and updated every 5 minutes for T and VZ.

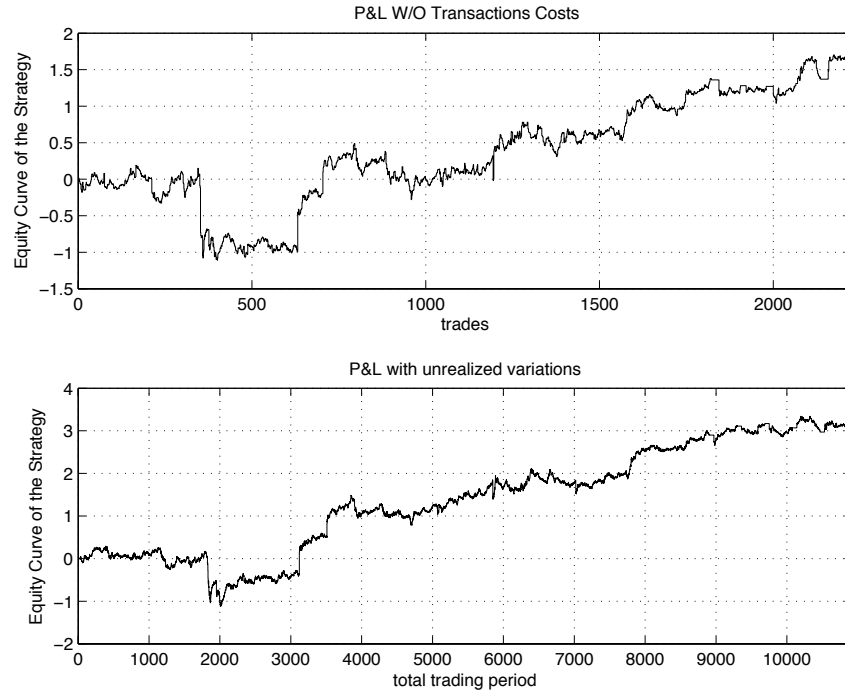


FIG. 32. P&L generated (out-of-the-sample) by the stochastic spread method for T and VZ.

Chapter 5

Conclusions and Future Work

In this thesis we have reviewed and discussed topics on quantitative finance, specially quantitative trading, both from a theoretical and empirical point of view.

In chapter 1 we have started with a **theoretical** background addressing some basic models for prices and spreads dynamics. Equipped with high-frequency data from the Dow Jones Industrial Average, we performed an **empirical analysis** on log-returns and log-spreads with different sample frequencies, namely 1-minute, 10-minutes, 30-minutes and 60-minutes. Concurring with some standard stylized facts we find that returns sampled at high frequencies presents near zero mean, decreasing kurtosis as function of sample size and at individual level (i.e: not aggregated such as an index) a power-law behavior does not hold. Also we report that the tail index is a decreasing function of sample frequency. We find that correlations are a function of time (figure 13), that intraday correlation exhibits a strong pattern due to human behavior (figure 10) and this measure is helpful to provide an appropriate measure of distance (eq. 38) to characterize the stock market and useful taxonomies arise to interpret a given economy or portfolio (figure 15).

In chapter 2 we address the question whether strategies from **technical analysis** are well defined in a scientific manner or not. We concluded that for any given technical rule in order to be well defined must pass the test of being a random or Markov time. That is, the variable τ which represents the signal to trigger a transaction must be \mathcal{F} -measurable. As a rule-of-thumb every technical rule involving averages are properly defined as Markov times. This is not the case of trends or patterns.

In chapter 3 we reviewed some basic forms of **market neutral strategies** from a theoretical point of view, such as the distance method, the

co-integration method, the stochastic spread method and the stochastic residual method. We have discussed both model specifications and parameter estimation including filtering techniques.

In chapter 4 we have implemented some of the models discussed in the previous chapters. This **implementation** involved minutely data from the Dow Jones Industrial Average from 01-Jun-2010 to 08-Nov-2010. Although the NYSE is very liquid, but some equities are not traded with our requested sampling frequency. The missing values are filled with the latest prices available.

The backtesting was done in the following way. First we split the sample into three quarters for in-sample calibration of model parameters and the rest for testing the profitability of the algorithm.

There are several aspect relegated in the previous implementation of various quantitative trading strategies which are vital if we intend to understand and asses the risk of a real quantitative trading system (quant system).

A “real” quant system would be composed as follows [38]:

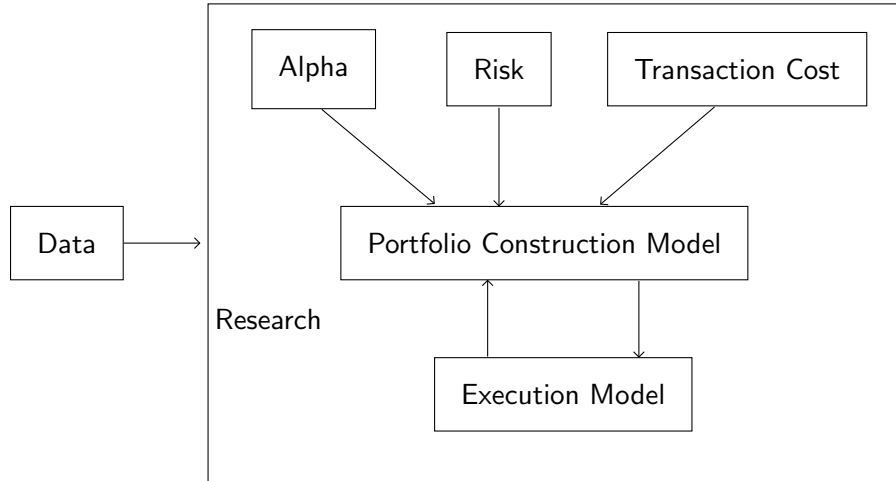


FIG. 33. A sketch of a quantitative trading system.

The so-called *alpha models* are composed by a variety of strategies but ultimately they can be cast in two big families: the theory driven models and the data driven models. This project was focused in the first family of models, specially in trend following or mean reversion models or price-related data, leaving aside factor models (fundamental data) yielding strategies such as value, growth and quality.

Risk models try to address questions like what size is appropriate to trade in each signal relative to the whole portfolio and to track the

size of the exposure in every trade. This subject was dealt only theoretically in section 9 of chapter 1 in the subsection *wealth dynamics*. Our approach was to find the optimal growth rate of a given portfolio. That is what fraction of our portfolio do an investor has to trade for maximizing its long term growth rate and minimize the risk of bankruptcy. This fraction has a name of its own, the Kelly criterion and it is due to Kelly and Shannon [45]. This was not implemented in our backtesting exercises in chapter 4.

Transaction cost models are typically divided in three classes. Commissions and fees are the most straight forward type to model. We can not trade with out a broker which charge clearing and settlement fees for executing our orders in any given market. The second type are the slippage costs. This are a little more tricky to model because we do not know a priori as the commissions. The slippage represents the difference of trigger prices and the execution prices. Here the latency (the speed to access to the market) is the key factor to model this types of costs. Finally, the last concern regarding to cost is the market impact. This kind of cost is typically associated with big market participants and market depth. When big orders are send to the market, this may is not deep enough to absorb the whole lot, and the price starts to move against the order. This difference is called market impact. The modeling of transaction cost is complex and participant dependent. Needless to say, the importance of modeling to asses correctly the potential profit of the strategy. This remains as a pending issue to future work.

Merging the above described models (alpha, risk and costs) results in a portfolio at a given time. The decision to allocate a certain amount to any strategy is based on a balancing of considerations of expected returns, risk and transaction costs. Typically, this allocation has two approaches. The first approach is based on some heuristics such as equal position weighting, equal risk weighting (size weighted as their inverse of volatilities), etc. The other approach has its roots on the so-called modern portfolio theory with the mean-variance optimization techniques and its variations, namely, Markowitz model, Black-Litterman, Factor portfolios, etc.

Finally, the last section of a quant system is the execution system. Order execution algorithms determine the way in which systematic execution of a portfolio is actually done. Typically two kinds of order are use to get the market: market orders and limit orders. The former one is submitted to the marketplace and is generally unconditional and it must be filled, that is it take whatever price prevails at the market at the time the order's turn to be executed. In contrast, the later type or order allows the trader to control the worst price at which he or she

is willing to transact, but the trader must accept that his order might not get executed at all or that only a part of it might be executed. A proper model to execute the orders in a quant system should have what is called limit order book. It is nothing but the collection of all available bid and offers for a given security. Usually this kind of data is quite expensive and some times is prohibitive to perform research at academic levels. The modeling of limit order books it is outside of the scope of this study.

The objective of this study was to build an introduction to quantitative trading aiming to clarify basic concepts, types of models and to backtest some strategies with real market high-frequency data.

As seen in chapter 1, specially in the empirical study, contrary to the mathematical laws of nature (proved extremely successful in the physical sciences), financial markets are not governed by invariance principles. They are driven by heterogeneous agents with different expectations about uncertainty which makes financial market alive, but not completely random. Hence, any attempt to model a small part of a complex reality should be dynamic and calibrated constantly to the changing market conditions.

As stated in the previous paragraphs, a real world quant system would involve a considerable amount of work devoted to the implementation and control of this strategies and special care in the execution part. But maybe this lies outside the interest of academia and pertains to the core of the financial industry.

References

- [1] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228. [39](#), [41](#)
- [2] Louis Bachelier. Théorie de la spéculation. *Annales Scientifiques de l'École Normale Supérieure*, pages 21–86, 1900. [8](#)
- [3] Oleg Bondarenko. Statistical arbitrage and securities prices. *The Review of Financial Studies*, 16:875–919. [3](#)
- [4] J.P. Bouchaud, Laurent Laloux, Pierre Cizeau, and Marc Potters. Random matrix theory. *Risk Magazine*, 12:193–197, March 1999. [30](#), [32](#)
- [5] Leo Breiman. *Probability*. Addison-Wesley, Reading, Massachusetts, first edition, 1968. [51](#)
- [6] Damiano Brigo, Antonio Dalessandro, Matthias Neugebauer, and Fares Triki. A stochastic processes toolkit for risk management. *Working Paper*. [65](#)
- [7] William Brock, Josef Lakonishok, and Blake LeBaron. Simple technical trading rules and the stochastic properties of stock returns. *The Journal of Finance*, 47. [4](#), [38](#), [39](#)
- [8] Michel M. Dacorogna, Ulrich A. Muller, Olivier V. Pictet, and Casper G. De Vries. Extremal forex returns in extremely large data sets. *Extremes*, 4:105–127. [21](#)
- [9] Emanuel Derman. Models. *Financial Analyst Journal*, page 47, February 2009. [3](#)
- [10] David A. Dickey and Said E. Said. Testing for units roots in autoregressive-moving average models of unknown order. *Biometrika*, pages 599–607, 1984. [65](#)
- [11] Binh Do, Robert Faff, and Kais Hamza. A new approach to modeling and estimation for pairs trading. *Unpublished paper*, page 31, May 2006. [5](#), [69](#)
- [12] Robert J. Elliott, John van der Hoek, and William P. Malcolm. Pairs trading. *Quantitative Finance*, 5:271–276, 2005. [5](#), [66](#)
- [13] Robert F. Engle and C. W. Granger. Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55:251–276, March 1987. [56](#), [58](#)
- [14] Eugene F. Fama. The behavior of stock-market prices. *The Journal of Business*, 38:34–105, January 1965. [20](#), [24](#)
- [15] Eugene F. Fama. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25:383–417, May 1970. [1](#)
- [16] Xavier Gabaix, Parameswaran Gopikrishnan, Vasiliki Plerou, and H. Eugene Stanley. A theory of power-law distributions in financial market fluctuations. *Nature*, 423, may 2003. [20](#)
- [17] Evan Gatev, William Goetzmann, and Geert Rouwwenhorst. Pairs trading: Performance of a relative value arbitrage rule. *Working Paper, Yale School of Management*, July 1999. [1](#), [4](#), [54](#)

- [18] Parameswaran Gopikrishnan, Vasiliki Plerou, Luis A. Nunes Amaral, Martin Meyer, and H. Eugene Stanley. Scaling of the distribution of fluctuations of financial market indices. *Phys. Rev. E*, 60:5305–5316, 1999. [20](#)
- [19] C. W. Granger. Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16:121–130, March 1981. [56](#)
- [20] James D. Hamilton. *Time Series Analysis*. Princeton University Press., Princeton, New Jersey, 1994. [59](#)
- [21] Michael J. Harrison and David M. Kreps. Martingales and arbitrage in multi-period securities markets. *Journal of Economic Theory*, pages 381–408, 1979. [3](#)
- [22] Daniel Herlemont. Pairs trading, convergence trading, cointegration. *YATS Finances & Technologies*, 2003. [5](#)
- [23] Bruce M. Hill. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3:1163–1174, September 1975. [22](#)
- [24] Robert Jarrow, Melvyn Teo, Yiu Kuen Tse, and Mitch Warachka. Statistical arbitrage and market efficiency: Enhanced theory, robust tests and further applications. *Johnson Graduate School of Management, Cornell University*, page 54, February 2005. [4](#)
- [25] Alireza Javaheri. *The Volatility Process*. PhD in Economics, Ecole de Mines de Paris, 2004. [34](#)
- [26] Soren Johansen. Statistical analysis of co-integration vectors. *Journal of Economic Dynamics and Control*, 12:231–254, January 1988. [56](#), [59](#)
- [27] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte. A new method for the non-linear transformation of means and covariances in filters and estimators. *IEEE Transactions on Automatic Control*, 45(3):477–482, March 2000. [36](#)
- [28] J. L. Kelly. A new interpretation of information rate. [41](#)
- [29] Fima C. Klebaner. *Introduction to Stochastic Calculus with Applications*. Imperial College Press, 57 Shelton Street Covent Garden London, 2005. [8](#)
- [30] Blake LeBaron. Do moving average trading rule results imply nonlinearities in foreign exchange markets. *The Journal of Finance*, 47. [4](#)
- [31] Blake LeBaron. *Technical trading rules and regime shift in foreign exchange, Advanced Trading Rules*. E. Acar and S. Satchell, Eds., Jordan Hill, Oxford, 1998. [4](#)
- [32] Benoit Mandelbrot and Howard M. Taylor. On the distribution of stock prices differences. *Operations Research*, 15:1057–1062, December 1967. [20](#)
- [33] R. N. Mantegna. Hierarchical structure in financial markets. *The European Physical Journal B*, June 1999. [27](#), [29](#)
- [34] R. N. Mantegna and H. E. Stanley. Scaling behavior in the dynamics of an economic index. *Nature*, 376:46–49, June 1995. [20](#), [21](#)
- [35] Robert C. Merton. Option pricing when underlying stock returns are discontinuous. *Massachusetts Institute of Technology (MIT), Sloan School of Management*, pages 787–75, 1975. [14](#)
- [36] Attilio Meucci. Review of statistical arbitrage, cointegration, and multivariate ornstein-unlenbeck. *Unpublished paper*, page 47, January 2010. [5](#), [67](#)
- [37] Franco Modigliani and Leah Modigliani. Risk-adjusted performance. *Journal of Portfolio Management*. [41](#)
- [38] Rishi K. Narang. *Inside the Black Box: the simple truth about quantitative trading*. John Wiley & Sons., Hoboken, New Jersey, 2009. [84](#)
- [39] Purnendu Nath. High frequency pairs trading with u.s. treasury securities: Risks and rewards for hedge funds. *Working Paper, London Business School*. [1](#), [4](#), [54](#)

- [40] Salih N. Neftci. Naive trading rules in financial markets and wiener-kolmogorov prediction theory: A study of technical analysis. *The Journal of Business*, 64:549–571, July 1991. [4](#), [45](#)
- [41] Marcelo Perlin. Evaluation of pairs trading strategy at the brazilian financial market. *Working Paper, Reading University*, November 2007. [1](#), [5](#)
- [42] Marcelo Perlin. M of a kind: A multivariate approach approach at pairs trading. *Working Paper, Reading University*, December 2007. [1](#), [4](#)
- [43] George Pólya. On the mean-value theorem corresponding to a given linear homogeneous differential equation. *American Mathematical Society*, pages 312–324, July 1922. [11](#)
- [44] R. Rockafellar, Stan Uryasev, and Michael Zabarankin. Generalized deviations in risk analysis. *Finance and Stochastics*, 10(1):51–74, 2006. [41](#)
- [45] Claude Elwood Shannon. A mathematical theory of communication. *Bell Systems Technical Journal*, 27:379–423,623–656, 1948. [41](#), [85](#)
- [46] R. H. Shumway and D. S. Stoffer. An approach to time series smoothing and forecasting using the em alogrithm. *Journal of Econometrics*, 16:121–130, March 1981. [66](#)
- [47] Adrian Trapletti, Alois Geyer, and Friedrich Leisch. Forecasting exchange rates using cointegration models and intra-day data. *Journal of Forecasting*, 21:151–166. [5](#)
- [48] Ganapathy Vidyamurthy. *Pairs Trading: Quantitative methods and analysis*. John Wiley & Sons., Hoboken, New Jersey, 2004. [4](#), [5](#)
- [49] Norbert Wiener. Differential space. *Journal of Mathematics and Physics*, pages 131–1974, July 1923. [8](#)

Glossary

A

arbitrage profiting from differences in price when the same security, currency or commodity is traded on two or more markets. For example an *arbitrageur* simultaneously buys one contract of gold in the New York market and sells one contract of gold in the Chicago market, locking in a profit because at that moment the price on the two markets is different.

C

co-integration is an econometric property of time series variables. If two or more series are themselves non-stationary, but a linear combination of them is stationary, then the series are said to be co-integrated.

K

Kelly criterion is a long-term progressive betting strategy which identifies the optimal bet in any situation, so as to maximize long-term bankroll growth.

M

mean reverting a theory suggesting that prices and returns eventually move back towards the mean or average. This mean or average can be the historical average of the price or return or another relevant average such as the growth in the economy or the average return of an industry.

Modigliani and Modigliani is part of a group of indicators called *risk-adjusted performance*. As opposed to the Sharpe ratio, this index aims to measure the returns of the portfolio adjusted to the risk of some benchmark. It is quoted in basis points.

P

pairs trading is a strategy that uses two highly correlated financial instruments whose price relationship has diverged outside of the historical range. In buying one and selling the other, the strategy aims to profit from the price reverting back to as the mean trend as the spread between the two converges.

P&L profit and loss statement. Summary of the revenues, costs, and expenses of a given strategy. Is the cumulative sum of the differences of the consecutive transactions of a strategy on a given period of time.

S

self-financing all trades are financed by selling or buying assets in the portfolio. No money is withdrawn or inserted after the initial forming of the portfolio.

Sharpe ratio average return, less the risk-free return, divided by the standard deviation of return. The ratio measures the relationship of reward to risk in an investment strategy. The higher the ratio the safer the strategy.

spread difference between the prices of two securities.

statistical arbitrage is an equity trading strategy that employs time series methods to identify relative mis-pricings between stocks.

T

technical analysis use charts or computer programs to identify and project price trends in a market, security, fund or future contract. Most of the analysis is done for short- or intermediate term. It is not concern with the financial position of the company but only to the demand and supply of any given security.

technical trading a method for trading securities according to rules from technical analysis.

Index

- alpha models, 84
- arbitrage
 - deterministic, 3
 - statistical, 1, 3
- backtesting, 84
- bootstrapping, 38
- Chapman-Kolmogorov equation, 33
- co-integration, 8, 53, 56
 - Engle-Grange method, 58
 - error correction representation, 61
 - filtering OLS estimators, 60
 - Johansen method, 59
- coherent measures of risk, 38
- empirical analysis, 18
 - correlation coefficient, 27
 - economical taxonomy, 26
 - Hill estimator, 21
 - probability distribution function, 19
 - scaling, 22
- ergodicity, 8
- expectation-maximization algorithm, 66
- filtering, 32
 - extended Kalman filter, 34
 - Kalman filter, 33
 - Kalman gain, 35
 - state space formulation, 65
 - unscented Kalman filter, 35
- hitting time distributions, 54
- Kelly criterion, 40, 85
- mean reversion, 16
- mean reverting, 5
- minimal-spanning tree, 28
- p-value, 38
- pairs trading, 4, 53
 - long position, 56
 - short position, 56
 - transaction cost matrix, 56
- performance indicators
 - Modigliani-Modigliani, 40
 - Sharpe ratio, 39
 - Sortino, 40
- portfolio's P&L, 55
- probability
 - finite Markov times, 50
 - Markov time, 7, 45
 - Markov times, 54
 - probability space, 6
 - sigma-algebra, 5
- quantitative trading system, 84
- random matrix theory, 29
- risk models, 84
- standard stylized, 83
- stationarity
 - strict, 7
 - weak, 7
- stochastic integration
 - Itô isometry, 17
 - Itô integral, 11
 - Itô's lemma, 13, 67
- stochastic process
 - Poisson process, 14
 - Wiener process, 8
 - Brownian motion, 8
 - geometric Brownian motion, 12
 - jump-diffusion, 15
 - Markov process, 10
 - multivariate spread dynamics, 67
 - Ornstein-Uhlenbeck, 16, 62
 - quadratic variation, 9
- stochastic residual spread, 69
- stochastic spread, 62

- t-statistic, 37
- technical analysis, 45
 - commodity channel index, 48
 - exponential moving average, 48
 - indicators, 4
 - moving average convergence
 - divergence, 49
 - moving average crossings
 - indicators, 46
 - oscillators, 48
 - rate of change, 48
 - relative strength index, 48
 - signal line, 49
 - stochastic index, 48
 - trend crossing methods, 47
- transaction cost models, 85
- ultrametric spaces, 28
- unit root testing, 65
- vector autoregressive process, 57

Appendix A

DJIA Companies

Symbol	Name
AA	Alcoa Inc.
AXP	American Express Company.
BA	Boeing Company (The).
BAC	Bank of America Corporation.
CAT	Caterpillar, Inc.
CSCO	Cisco Systems, Inc.
CVX	Chevron Corporation.
DD	E.I. du Pont de Nemours.
DIS	Walt Disney Company (The).
GE	General Electric Company.
HD	Home Depot, Inc. (The).
HPQ	Hewlett-Packard Company.
IBM	International Business Machines.
INTC	Intel Corporation.
JNJ	Johnson & Johnson.
JPM	JP Morgan Chase & Co.
KFT	Kraft Foods Inc.
KO	Coca-Cola Company (The).
MCD	McDonald's Corporation.
MMM	3M Company .
MRK	Merck & Company, Inc.
MSFT	Microsoft Corporation.
PFE	Pfizer, Inc.
PG	Procter & Gamble Company (The).
T	AT&T Inc.
TRV	The Travelers Companies, Inc.
UTX	United Technologies Corporation.
VZ	Verizon Communications Inc.
WMT	Wal-Mart Stores, Inc.
XOM	Exxon Mobil Corporation.



Campus de Bellaterra, Edifici C
08193 Bellaterra, Spain
Tel.: +34 93 581 10 81
Fax: +34 93 581 22 02
crm@crm.cat
www.crm.cat